Michell, Elene.

Teaching values in new-type history tests

VALUES IN NEW-TYPE HISTORY TESTS

MICHELL

[redacted]

Michell, Elene.

Teaching values in new-type history tests

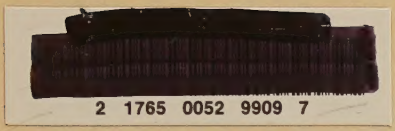
Class No. 907 Date rec'd 1/21/31

Book No. M623 Donor

Accession No. 24434 Fund App.

This book should be returned at the time indicated on date due slip; otherwise the Librarian is required to collect a fine for overtime.

13–2147


2 1765 0052 9909 7

TEACHING VALUES IN NEW-TYPE HISTORY TESTS

By Elene Michell, Ed.D.
Associate Professor of Social Science
State Teachers College
San Francisco
California

WORLD BOOK COMPANY
Yonkers-on-Hudson, New York
1930

M623

WORLD BOOK COMPANY

THE HOUSE OF APPLIED KNOWLEDGE

Established 1905 by Caspar W. Hodgson

YONKERS-ON-HUDSON, NEW YORK

2126 PRAIRIE AVENUE, CHICAGO

ALSO BOSTON : ATLANTA : DALLAS

SAN FRANCISCO : PORTLAND : MANILA

THERE is an old story of a man shingling a roof in a heavy fog, who suddenly discovered that he had worked his way off the roof and had shingled out three feet on the fog. Too much of this kind of thing has been done in teaching the social subjects. In history especially, the good talker and the good guesser have often been allowed to "nail shingles on the fog" — to deliver themselves of opinions worthless because not based on a sufficient fund of facts. *Teaching Values in New-Type History Tests* shows how the teaching of history can be strengthened by making adequate provision for the acquisition of information. At the same time it shows how the informational aspect of the subject can be relegated to its proper place, allowing time and releasing energy for thought and reasoning, without which the learning of facts is vain.

24434

MTVNTHT-1

Copyright 1930 by World Book Company
Copyright in Great Britain
All rights reserved

PRINTED IN U.S.A.

PREFACE

THIS material is designed to aid the teacher of the social studies in the upper grades of the elementary school or the junior or senior high school in the use and construction of objective tests. Two underlying contentions will be found in these pages: The first is the insistence that testing is a part of method in teaching a subject, and the second is the claim that the acquisition of some information is a definite, even though subordinate, objective in the teaching of history. Unless the first contention is accepted and acted upon by the teachers of history and testing becomes more than a matter of devising questions from a given text, instructors will be forced by the momentum of the testing movement to modify the material they teach in order to prepare for tests developed by experts in measurement. The matter of content and objectives in tests as well as in subject matter is primarily the concern of those skilled in methods of teaching a given subject, and it is highly desirable that teachers of history determine test content through the purposeful adaptation and use of new examination forms.

The second contention has arisen out of the difference of opinion between those who assert that their first concern is the subject matter of history and those who assert that their chief concern is the use of progressive methods of teaching. By means of an analysis of thinking based on historical information and the development of methods in testing that place the responsibility for memory work in the field of pupil activity, it can be shown that there is no conflict between the stressing of historical accuracy and the use of progressive methods in teaching.

The term "history" has been used in its broadest sense as a record of man's progress contributing to a better understanding of our own times. Material that is often listed as

civics, economics, and geography is treated as history. In order to be specific, illustrations have been drawn from the teaching of United States history in high school, for there is much similarity between the principles of method for the social studies in the upper elementary grades and the junior and senior high school.

ACKNOWLEDGMENTS

DIRECT acknowledgment is given in the following pages to writers who have been quoted. Such citations, however, cannot include all those whose writings and instruction have contributed to the theory here set forth. Recognition is due certain progressive educators and writers on method in the social studies as well as those who have conducted surveys and experiments in this field.

Gratitude is accorded the principals, history teachers, and students of the four high schools participating in the experiment with tests reported in Chapter VII. The high schools are in Watertown, Massachusetts; Redondo Beach, California; and Los Angeles (Belmont High School), California. "Z" school, which prefers to remain unnamed, is in Massachusetts. The success of the experiment owes much to certain professors in the Graduate School of Education in Harvard University for suggestions and advice. Appreciation is due Miss Marien M. Keyes and Dr. J. S. Orleans, Test Editor of the World Book Company, both of whom read the manuscript and made valuable suggestions.

The following publishers have permitted quotations from books: D. Appleton & Co.; D. C. Heath & Co.; Houghton Mifflin Company; Randolph-Macon Woman's College; Scott, Foresman & Co.; Teachers College, Columbia University; Macmillan Company; National Council for the Social Studies; and World Book Company.

Permission has been given to quote from the following periodicals: *Harvard Graduates' Magazine, Historical Outlook, Journal of Applied Psychology, Journal of Educational Research, Journal of Education, Journal of Educational Psychology, Psychological Bulletin, Atlantic Monthly, Nation, San Francisco News.*

E. M.

CONTENTS

CHAPTER PAGE

I. The History Teacher's Dilemma 1

The difficulty of securing constructive thinking in history — Need for historical perspective in the problem approach — Temptation to loose discussion — The dilemma presented by progressive methods — The problem method supplemented by objective tests — Why test development in the social studies has been slow — Maintaining individuality of thinking through tests constructed by the teacher — Ways in which tests aid the teacher — Need for tests of information

II. Reflective Thinking in Relation to Studying History . 11

Effective attitudes and behavior the ultimate aim — Attitudes the result of specific experiences — Fallacies related to training for attitudes — The connection between factual memory and reasoning — The nature of reflective thinking — Need for conditioning thinking in specific fields — Organized associations essential to clear thinking — Facts a means, not an end — Linking of attitude, information, and reflection — Training needed for voters — Unthinking emotionalism and the mob mind

III. Characteristics of New-Type Tests Valuable in the Social Studies 30

The purpose of examinations — Distinctive functions of the essay-response and new-type examinations — Objectivity of new-type examinations — Time factor involved in improving and correcting essay-response tests — Characteristics of new-type tests of value in the social studies — Comprehensiveness — Evidence of comprehensiveness noted in certain studies — Reliability increased by comprehensiveness — Evidence of reliability in new-type tests — Evidence of lack of reliability in old-type tests — Increased validity of objective tests — Historical information as distinct from skill in expression measured by objective tests — Measurement of higher objectives in a social studies test demanded by validity

IV. Using Tests to Improve Instruction 41

DIAGNOSIS OF LEARNING DIFFICULTIES. Value and availability of reading comprehension tests — Other diagnostic tests — The Working Skills Test — Means of measurement but not of preliminary diagnosis furnished by certain published tests — Inventory tests

TESTING TO INDIVIDUALIZE INSTRUCTION. Seeing the student as an individual through use of diagnostic tests — Improving study habits — Improved teaching technique and time saving

CHAPTER PAGE

effected by tests that provide check of individual reading records and exercises for problem study — Use of exercises for problem study illustrated

RAISING THE STUDENT'S STANDARDS. Student's perception of goal sharpened by tests — Need for historical accuracy emphasized — Self-measurement motivated — Effort rewarded

PROVIDING A MORE EFFICIENT MEASURE OF ACHIEVEMENT. Discrimination of values — Improved marking system — Superiority of rank method — Percentile rank

V. NEW EXAMINATION FORMS ADAPTED TO HISTORY . . 63

Discussion of various objective forms — Recall form — Completion form — Multiple-response — Recognition — Best-answer — Topical-selection — True-false — Advantages, weaknesses, and variations in true-false form — Difficulties in constructing true-false tests — Distinctive qualities in question forms — Time sequence — Reliability of various forms — Conclusions of Iowa study — Scoring objective forms — The "rights-minus-wrongs" controversy — Unweighted scores satisfactory — Deliberation and the time limit — Other ways of rating ability — Items possible in a given time — Reasoning tests and the single-answer response — Confusion in use of term *reasoning* — Cooperative criticism needed for tests of ideals — Subjective judgments in multiple-response form — Illustrations from published and unpublished tests

VI. CONSTRUCTING AND ADMINISTERING AN INFORMAL TEST . 93

RELATING THE TESTS TO THE TEACHING PLAN. The work sheet — Emphasis on study, not on the test mark — Need for practice tests — Discussion of test results with class — Explanation of simple statistical terms — Small amount of time given to testing

ASSEMBLING THE MATERIAL FOR A TEST. The basis for selection — Apportioning the material — Checking the preliminary form — Apportioning emphasis to various epochs — Adapting question form to content in the true-false, multiple-choice, and topical-selection forms — Qualities of a good question — Difficulties to avoid in constructing a test: poor selection, ambiguity, faulty form

ADMINISTERING AND SCORING TESTS. Directions to coöperating teachers and to students — Completed form — Mimeographing — Margins and spacing — Safeguarding the honest student — Need for care in scoring — Directions for scoring — Student help in scoring — Form subordinate to content

VII. AN EXPERIMENT WITH UNIFORM TESTS IN FOUR HIGH SCHOOLS 121

OF WHAT VALUES ARE UNIFORM TESTS TO A DEPARTMENT? Uniform tests in sections studying same period of history —

CHAPTER PAGE

Equalizing minimum requirements — Evidence of varying standards in a department

WHAT INFLUENCE SHOULD THE RATING ON AN OBJECTIVE TEST HAVE ON THE TEACHER'S MARK? Evidence of inconsistency in marking — Cases in which marking was reasonably consistent — Mechanical marking not desired

CAN THE TEACHER BE JUDGED FROM THE CLASS MEDIAN? Characteristic level shown by scores for series of tests — Measuring through varied exercises

WHAT ACHIEVEMENT CAN BE EXPECTED IN HISTORY FROM THE STUDENT OF NORMAL INTELLIGENCE? Scores in series of objective tests in history arranged in scale on basis of IQ and of average test score — Conclusions concerning attainment that can be expected

IS THERE A COMMON BASIS FOR MEASUREMENT IN HISTORY? Similarity in median scores in three of the four schools studied — Conclusions

APPENDIX: ILLUSTRATIONS OF TESTS USED IN AN EXPERIMENT IN FOUR HIGH SCHOOLS 143

Note to teachers participating in the experiment — Test 2A: Period of Exploration and Colonization — Test 2B: Alternate form for A — Test 3A: Winning Independence — Test 3B: Alternate form for A — Test 5A: Colonization and Independence — Scoring keys for various test forms

BIBLIOGRAPHY 169

INDEX 177

TEACHING VALUES IN NEW TYPE HISTORY TESTS

CHAPTER ONE

The History Teacher's Dilemma

History was long considered one of the easiest subjects to teach because instruction consisted merely in helping the student memorize the text. The tacit acceptance of this fact was indicated by the remark of a principal concerning a weak teacher: "She cannot teach anything but history, but she can 'get by' with that since she needs only to hear the children recite what is in the book." It is becoming increasingly apparent that the very ease of this mock procedure makes history one of the danger spots in the curriculum. A high degree of teaching skill is required if the student is to gain stimulation to constructive thinking in civic affairs as well as general information. The effort to induce original thought in this field presents a unique difficulty, since the student's opinions are of no value unless they are based upon a knowledge of contributing facts. Even worse, loose thinking in the adult citizen is at once a characteristic of American democracy and a menace to its development.

Modern educational methods demand learning through the solving of problems. Only thus is individual thinking challenged and true learning achieved. In the social studies the situation is very different from that obtaining in the physical laboratory, where the various substances needed for an experiment may be assembled, the demonstration given, and the student left with the materials to repeat the experiment or to solve related problems. In the history class there are no materials that can be presented unless there is

some previous knowledge of social and political situations such as is implied in geographical and historical familiarity with men and nations. Without that, how can students grasp the problem that a teacher would like to make the starting point of study?

The history teacher needs the skill of a juggler handling a two-edged sword. For though a history room include a museum and a reference library, analytical and constructive thinking is impossible unless the student possess some understanding of the sequence and interdependence of past events. And, just as surely, though the student may be able to tell off countless dates and genealogies, there has been no "study" of history if there is no understanding of causal relations. Probably each subject in the school program presents difficulties both unique and complex, but the teacher of history feels that no other subject makes such a contradictory demand as this — the requirement for an initial "problem attack" and the impossibility of understanding the problem without a knowledge of underlying facts.

Furthermore, both the problem method and the necessity for motivation tempt a teacher to permit loose discussion. Some freedom is essential if the deadly formalism of the old method of teaching history is to be eliminated, but there are dangers associated with the reaction from empty memorization. The tendency toward loose thinking and the formation of snap judgments has come in for much criticism in discussions of "progressive" methods. In summarizing the conclusions of a nation-wide survey, Dawson gave this warning:

> Here may be mentioned one of the dangers of the present teaching of the social studies in some quarters. This danger lies in the confusion of thought between opinions of ignorant people and those of trained minds, on the one hand, and, on the other, the difference between opinions and mere emotions or impressions. Closely

related with this confusion is the apparent belief on the part of some teachers and college professors that subjective unsupported opinions are rather more important than practical men might generally consider them. Out of this situation sometimes grows a disposition on the part of teachers to permit pupils to air what are called opinions in unprofitable class discussion.[1]

In an article published about the same time Munro discussed the same tendency:

We should not confuse discussion with thought. So long as talking is so much easier than thinking, we shall have far more of the one than of the other when we give them equal opportunity. The socialized recitation, as I have observed it, very often deteriorates into a babbling match with the maximum of talk and the minimum of reflection. The most conspicuous intellectual vice of the American youth is his readiness to express a definite opinion on any subject at any time, and to do this without a knowledge of the facts or of the forces which are behind facts. I have found, for example, that if I ask a group of college freshmen to discuss the wisdom and expediency of compulsory voting there is hardly a youth in the room who does not stand prepared to give me the pros and cons right off the reel. But if I ask whether compulsory voting has ever been tried in any country, and with what results, I find that not one of them has had the idea of approaching the matter from that point of view. To get the facts is irksome and takes industry; it is much easier to confine the instruction to "opinion-forming," and thus let the gold-dust twins do the work.[2]

The history teacher faces a dilemma. Undoubtedly as much harm has come through the type of history teaching that overemphasizes factual material as through the other extreme that encourages loose thinking and the expression of unsupported opinion. Yet it is absurd to ignore the value either of facts or of independent thinking merely because there are difficulties in the way of presenting both parts of

[1] Edgar Dawson, "Report of the History Inquiry," in *Historical Outlook*, Vol. XV, No. 6, page 263; June, 1924.

[2] William Bennett Munro, "The Gold-Dust Twins in Education," in *Journal of Education*, Vol. CII, No. 7, page 174; September, 1925.

this learning process. An interest in the study of history is not destroyed by the recognition that accuracy of statement and the citation of evidence are essential in dealing with historical facts. The dislike for history usually arises because students see in it nothing but memorization, and, too frequently, memorization of military and political events only. Of course, sadly enough, a few students find in this their only reason for liking the subject. As one naïvely said, "You learn what is in the book and you are through; you do not have to waste your time wondering if you are right." While there is no condemnation severe enough for such teaching, it must be admitted that the other extreme is just as injurious to the student. That point of view is exemplified by the one who said, "I like civics because you do not have to learn anything; you just read the papers and say what you think."

The emphasis that progressive educators have put upon the development of initiative and upon the necessity for proper motivation has to a great extent freed the study of history from formalism and memorization. But without a recognition that such freedom increases the difficulty of teaching the subject, there is a tendency to let liberty become license. No one knows as well as does the conscientious instructor of junior and senior high school classes the difficulties that are presented by the use of progressive methods. The subject had, perhaps, better never be taught than become a mere repetition of the text; but there is certainly a grave danger from the other extreme of free discussion regardless of evidence and from the acceptance of aims that concentrate effort upon the production of desultory object work that has no continuing purpose or unity in developing the subject. A city supervisor once said to the members of the social science department of a high school, "If you want to be appreciated, do work that shows. Just now we are sending

all our visitors to S——— High School to see the model of a medieval castle that has been built there." A criticism of the supervisor's comment is no condemnation of the project, but it is a plea for balance. Progressive methods are capable of abuse as well as are the more conventional forms of teaching; sometimes such methods seem to play into the hands of the superficial teacher.

While conceding the outstanding contribution of the progressive school in maintaining that learning must be motivated through real interest, it cannot be denied that the understanding of problems is impossible unless the related facts are known. The necessary facts are of course dependent upon the problems selected; there is no need to study all problems or all facts in our history. It is better to select certain outstanding problems as typical ones, problems relating to our political organization or present issues, or problems entailing a knowledge of facts needed for the solution of still other problems — such, for instance, as the problem of Civil War finances, entailing a knowledge of currency, reserves, refunding, and taxation — facts and relationships essential for present-day issues as well as for those of the Civil War. The relation between problems and facts can be seen in the following illustration based upon a problem that is significant not only for the past but for the future:

What is the policy of the United States in regard to imperialism?

Which of the accessions of territory that have taken place since the accession of Alaska in 1867 have been direct purchases? Which have not?

What political events have been associated with the various accessions? Were the accessions due to political, military, commercial, or social needs?

What statements does the Constitution make relative to the government of territories? Under what clauses are subject colonies acquired and governed?

What Constitutional issues concerning citizenship, tariff reciprocity, and local control have found echo in political and legislative happenings in the United States?

Have these holdings affected our international policy?

What events can be considered illustrative of this fact?

What international organizations have we joined as a result of insular holdings?

What semiofficial protectorates are we holding?

What evidence can you cite to show that such policy has been advantageous, or otherwise, to the nation?

The understanding of this problem situation demands a comprehension of many basic facts in American history. To read over a statement of the actual situation does not indicate, however, that the student has assimilated the knowledge; but to be able to recall facts for the purpose of answering questions related to a problem is some evidence of learning.

Objective tests aid in solving the difficulty of combining the problem method with factual learning through furnishing a teaching device that puts upon the students the responsibility for learning certain basic facts and allows the teacher freer use of the recitation hour for study guidance and discussion. Secure in the realization that a series of comprehensive tests will check the student's possession of basic knowledge, the teacher may devote more time to the study of special problems. If the student knows that a ten-minute test may occur any day, and that there will be summary tests at the close of each unit as well as at the end of the term, he will face the continuous need for knowing the major points related to the class discussion.

It is perhaps fortunate that the development of standardized tests in the social studies has been slow. This is a situation due not only to the indefinite and changing status of both content and method but also to the fact that the controversial nature of political interpretations makes them

unsuitable for the single, brief-answer requirement of the objective response. Undoubtedly, too, the opposition of the " progressive " school to teaching through factual learning has restricted experimentation. Ultimately testing in the social studies will gain from this conservatism, for the actual teachers are thus given an opportunity to develop their own testing exercises according to the material taught instead of having the subject content determined by the need to see their classes rate high in a series of standardized tests.

The suggestion of standardized tests raises many problems in the social studies. While historical accuracy is essential, the interpretations put upon the past in order that we may better understand our own times express a constantly shifting emphasis, and in the nature of progress such changes must continue. If standard tests tend to crystallize interpretations, they may become a hindrance to progress. Of this we need to beware since the schools are already accused of instructing youth with doctrines that were current in the school days of the now mature instructors. The history of standardized tests, however, indicates little danger from this source because of the fact that the life of the usual standardized test is not sufficiently long. New tests in each subject are continually appearing, representing in general an improvement over existing tests in the same subjects. Tests are standardized and given to the public, merely to be improved and supplanted by fresh standardizations. There are other problems nevertheless, vitally concerned with the development of individual thinking, that must be considered in this connection. Because of the dangers of superimposed content, it is essential that progressive classroom teachers experiment with the new examination forms and guide their development as an instrument of better teaching — for, after all, *teaching and not measurement is the dominant purpose of examination.*

Objective testing may aid the teacher in several ways — as a preliminary diagnosis of learning difficulties, as a part of the teaching process, and as a comprehensive measure of achievement. In the consideration of these different values, emphasis will be placed upon the use of the new-form examination as a teaching device that stimulates the students. Testing for measurement has become common practice in educational systems, but testing as an aid to instruction has not received the attention that it deserves. Within the past few years a number of tests in history and civics have been published. For the most part they attempt to measure completed study or reflect the old, iniquitous procedure: "Study these questions for tomorrow and write the answers in class." Nothing can be further from the proper study of history than such an attitude, and yet this is the basis of one book on the use of tests in history. It is essential that testing and test construction in the social studies be considered a problem of teaching method. The difficulties of combining factual learning with the "progressive" method of teaching history will therefore be discussed first. Later chapters will treat of the adaptation of certain new examination forms to history, the construction of tests, and the results of an experiment in using such tests.

In considering the new-type examination in history, the questions often arise: Why stress tests of information? Why not use reasoning tests? Later chapters will answer these questions in detail. The following is a brief summary of the reasons given later: (1) Because information tests emphasize an element frequently neglected in the discussion method; (2) because such tests make for the elimination of factual drill from the recitation and thereby enable the teacher to emphasize other types of learning that require the aid of the skilled instructor; (3) because the subjective factor in original opinions and interpretative thinking is not so easily

covered by the objective test that allows only one correct answer, and hence the beginner in test construction should experiment first with the simpler medium of facts; and (4) because the fact test includes much more than is commonly attributed to it; the psychological process of reflective thinking is so closely associated with the complex recall of information that not only is reasoning dependent upon information but much that is commonly called reasoning is complex association or remembered interpretation.

The fourth point involves an analysis of the psychology of reflective thinking in relation to the study of history and has been considered important enough for discussion in a separate chapter (Chapter II), which the reader may omit if he desires without losing the continuity of the material relating to test use and construction.

Further development in objective testing will undoubtedly include other than factual material. New forms better adapted to reasoning will be devised, and existing forms will be modified to express individual judgments better. But the single-answer question is a device so valuable for economy in learning and so fraught with danger in teaching that it is imperative for the one who begins experimentation to recognize the essentially factual nature of evidence as distinguished from the essentially free nature of individual opinion.

Summary. History is one of the most difficult subjects to teach because modern methods demand the initial presentation of a problem, and yet the student's conception of the problem depends upon his possession of fundamental information and historical perspective. There is a tendency among some adherents of the " progressive " school to detract from the value of their own method through failing to concede the relation between the possession of historical information and the determination of historical judgments. Objective tests aid in solving this difficulty through furnishing a teaching

device that puts upon the student the responsibility for learning certain basic facts and allows the teacher freer use of the recitation hour for study guidance and discussion. There are reasons fundamentally sound for the conservatism attending the development of standardized tests in the social studies. Better tests will come from the experimentation of actual teachers of history in the new-type examinations. The beginner in test construction will do well to master the mechanics of the new-type question through experimentation with factual material.

CHAPTER TWO

Reflective Thinking in Relation to Studying History

In recent years the question has frequently been raised as to why facts should be taught in history and the other social studies. Since the purpose of such teaching is to make for better civic behavior, it is urged that the direct inculcation of desired attitudes should be undertaken without the risk of deadening the pupil's responses through memorization of facts. Such arguments ignore the psychology of learning and resort to fallacious reasoning. An analysis of the way in which attitudes are related to behavior indicates the futility of the attempted short cut.

The ultimate aim. That right behavior determined by effective attitudes and ideals is the aim of instruction in history is a statement of such evident truth that it leaves little room for dissent. If the history teacher could know that her work would result in attitudes of social responsibility, of directed purpose, of open-minded questioning for truth, she would be willing to sacrifice any requirements about the mastery of subject matter. But attitudes of right behavior can no more be poured ready-mixed into the child's mind through a process of telling than can ethical conduct be induced through a study of a catechism. The great problem of education — the effort to influence future behavior through fusing ideals with accumulated knowledge — exists in heightened form in the social studies. The existence of two such diverse factors in learning as the mastery of accumulated knowledge and the development of initiative has always been recognized. The early academic ideal was concerned almost entirely with transmitting information, while recent "progressive" schools center interest in the child's immediate behavior and expect knowledge to result

from increased initiative. The development of attitudes leading to wise behavior demands a blending of these two factors, and this blending is in turn dependent upon giving the youth activities that make for the connection of the factors. Actual experiences in a school democracy where history and civics become tools for the school citizen aid, in a unique manner, in the solution of this problem. The students who elect the "popular" fellow for student-body president and find him lacking in the wisdom or firmness needed to deal with offenders in the school democracy have learned that the easy-going qualities contributing to popularity may be anything but desirable in the executive official upon whose firmness the administration of justice depends. But whether the school democracy exists, or whether the attempt is made to educate in civic ideals only through vicarious experience such as the study of history presents, there must be recognition of the essential necessity for the exercise of reflective thinking if behavior is to be wise.

Attitudes do not exist in isolation. Attitudes and ideals are fundamental in determining behavior, but they are futile and uncertain if unrelated to reason and contributing evidence. To develop initiative and to ignore historical fact is a dangerous method of training youth. When an alert consciousness of self exists without the acceptance of social conventions or a knowledge of the significance of race traditions, the individual's value to himself and to society is questionable. The leaders of the French Revolution served humanity through their idealism, but they failed their country in its immediate need because their initiative was not combined with a wisdom that conceived the necessity for a stability built upon accepted traditions. The idealistic attitudes of the leaders were detached from experience, and they held as valueless any evidence contrary to their desires. A right social attitude cannot be considered as an isolated

entity, but is composed of various elements, such as idealism, an understanding of social organization, and practical experience.

Teaching that results in an attitude of indifference toward the element of evidence in political affairs will not result in an alert electorate. The forces that determine behavior are a complex interweaving of myriad tendencies, directed by various circumstances toward specific responses. Ideals are of primary importance in forming right attitudes, but to make the idealistic attitude effective there must be the ability to blend the ideal with the raw material of the actual situation. Idealistic revolutionists have failed not because they have lacked an attitude of philosophical idealism but because this attitude was detached from reality and associated with "wishful" thinking. Right attitudes are a composite of an indefinite number of psychologically conditioned responses, and will be socially valuable only as they are linked with a tendency to react to specific life situations.

Attitudes result from specific experiences. The instructor cannot afford to forget for a moment the fact, which modern psychology has so emphatically demonstrated, that "telling is not teaching." It is part of the innate character of learning that attitudes result from definite experiences that have more or less profoundly affected consciousness. It is as futile to expect social responsibility to result from studying about its desirability as to assume that honesty will result from enforced attendance at Sunday school. Even more the teacher must remember that an attitude is a coloring of individual consciousness and may develop very naturally into opposition to the desired concept instead of an acceptance of it, as when an attitude opposed to law enforcement becomes associated with a supposed limitation of personal liberty. Learning that is to be effective in adult life will result only from specific situations that have some common

thread linking the school experience with the type of activity needed by the adult. Any normal high school youth will express an attitude of respect for law, but that does not prevent this same youth from speeding on the highway or at intersections if he thinks he can "get away with it." But if the boy has served as a traffic aid at intersections near the school and witnessed the disastrous results of recklessness, he has learned, in a way not easily forgotten, the need to drive carefully.

Two fallacious deductions. The recognition that attitudes are of more significance in learning than factual memorization has been so revolutionary in its upheaval of old forms of teaching that two erroneous deductions have crept into certain educational theories. The first of these fallacies is that the acquisition of information is useless, and the second is that the attainment of right attitudes displaces any other form of learning. Granting without hesitation the force of the premise that the learning of isolated bits of information is useless, it does not therefore follow that information is useless, nor that the development of right attitudes supplants any other type of learning. It is the fact of isolation, the lack of meaningful associations, that makes memorization so profitless. And it is just as true that attitudes will not exist and function unless they are intertwined with the total mass of activity and based upon concrete experience. The American revolutionists were more effective in their attitude of opposition to tyranny than were the revolutionists of France because Washington and his colleagues had a rich fund of experience in the administration of local affairs to draw upon.

Educators cannot afford to ignore the relation of attitudes to experience. In its emphasis on the force of experience in learning situations, psychology is not presenting a new truth. There is abundant evidence in proverbs, such as "bumping

your own head," "holding the umbrella yourself," and other bits of folk-wisdom to show that the value of experience has been long appreciated; but psychology has brought new facts about experience to our attention. Two of these truths are of prime importance to educators: (1) that instruction must provide opportunity for experience; and (2) that, since experience is always concrete, it must be connected with specific responses which will carry value in adult life.

In the field of history few experiences or specific reactions result merely from talking about civic virtues, nor are attitudes likely to be transferred to adult life unless there is some common element of specific information or experience that is significant for the present life of the youth as well as for his future.

Experience makes memories permanent and fruitful. The young student accepts quite readily the emotional attitude that makes the Civil War a conflict carried on to free slaves. Not so easily does he see the underlying conflict of states' rights as an issue that is still unsettled and may yet threaten the unity of the Republic and the principle of majority rule. But when the issue of Federal control is connected with the Prohibition Amendment as well as with secession, a situation is presented that has meaning for the high school student of today. Through his own experience he may approach with more understanding principles of government linking the past with the present. While establishing such a relation is requisite in teaching history, the alert teacher will also connect with the idea of Federal control other significant conflicts over the question and relevant facts concerning the amendment of the Constitution. Unless the learning experience is connected with principles of government that are of continuing importance, learning has no common element to carry into situations of adult citizenship, and the discussion of states' rights versus Federal control will mean

no more to the student's future activity than the solution of a crossword puzzle. In such cases, reflective thinking is dependent upon immediate experience to provide problem situations, just as it is also dependent upon knowledge about conditions radiating from this experience to provide a basis for the solution of the problem.

Effective learning must be based upon specific responses. The fallacy that information is a negligible factor in education arises from a partial understanding of the truth that modern psychology has made so plain — the fact that learning is vacuous if divorced from personal experience; it does not become dynamic. Our democracy needs a type of learning in regard to government that will not only continue the highest ideals of government but will also result in practical activities in connection with them. The mere acquisition of information about the history and structure of government will not result in civic usefulness. Something must happen to make this knowledge effective if improved citizenship is to result. This something is the recognition by each voter of the significance of individual responsibility. The mental activity — or attitude — that will transform memory of historical events into improved citizenship is dependent upon experiences and knowledge that will widen the world of individual consciousness to effect a unity with the social whole. And further, the breadth of this social understanding is dependent upon the scope of specific information and personal, though perhaps vicarious, experiences that make attitudes a vital force in behavior.

The connection between factual memory and reasoning. An essential factor in developing effective social attitudes is reason. Regardless of the force that arouses mental activity, it is evident that reflective thinking cannot go on without some sort of informational background. Dewey says:

Thinking cannot of course go on in a vacuum; suggestion and inference can occur only upon a basis of information as to matters of fact.[1]

In discussing the same problem in college teaching, Wood comments as follows:

There is not as much opposition between "information" and "reasoning" as some teachers would have us believe. We remember by thinking, and we think by remembering facts. Those who declaim against tests of "mere facts acquired and remembered" should realize that we cannot think without facts.[2]

For the moment, our analysis is concerned not so much with the initial impulse of critical thinking as with the substance upon which this structure of thinking is built. The close relation that exists between thinking or reasoning and facts has been very clearly stated by Miller:

The power of thinking cannot be trained in the abstract, or in isolation from the process of acquisition of knowledge. Possession of knowledge in any given field is the inevitable corollary of past association and plays a vital part in thinking. Success in finding the relevant associates, the elements necessary to the solution of the problem in thought, is contingent upon their number and their organization. One's stock of definite and well-organized knowledge within the field of the problem is an essential factor in the judgment of what is relevant and what is irrelevant to the solution.

Who is the most likely to be successful in thinking through problems in economics, in history, in biology, in psychology, in engineering, etc.? The man who knows the most about these subjects and who is trained in the technique characteristic of each. Thinking power is not an abstract and general power of the mind to be applied equally well in all sorts of situations. It is rather a function of some larger whole, varying with the degree of development of that larger whole. The larger whole includes special knowledge of fact and special training in the technique of the subject. The good thinker in mathematics may be a very poor thinker in

[1] John Dewey, *How We Think*, page 52. D. C. Heath & Co., Boston; 1910.

[2] Ben D. Wood, *Measurement in Higher Education*, page 162. World Book Company, Yonkers-on-Hudson, New York; 1923.

economics or sociology, and vice versa. The habit of care in the examination of data, in the analysis of a situation, etc., may be carried over from one department to the other, but the special knowledge and the training in the special technique of one may be of little or no use in the other. The thinking process falls within systems of organized fact, as well as being a factor in the organization of material.

If these things are so, we delude ourselves when we think of such a thing as training children to think apart from the process of building up a body of knowledge. . . . If we are to train children of any age to think, one of the factors in this process is the building up of a system of definite and exact knowledge of facts within the sphere in which the problems of thought are to arise.

A great deal of loose teaching which does not show results in the definite knowledge of important facts and truths mastered by the pupils is unjustifiably excused on the ground that the chief stress has been put upon ability to think. The fact is, there is no ability to think in geometry without knowing the facts of geometry, in geography without knowing the facts of geography, etc.[1]

The nature of reflective thinking. Dewey has analyzed reflective thinking into five steps:

Upon examination, each instance reveals, more or less clearly, five logically distinct steps: (1) a felt difficulty; (2) its location and definition; (3) suggestion of possible solution; (4) development by reasoning of the bearings of the suggestion; (5) further observation and experiment leading to its acceptance or rejection; that is, the conclusion of belief or disbelief.[2]

The importance of the critical or selective agency in reflection is also emphasized. Dewey says that this implies:

(*a*) a certain store of experience and facts from which suggestions proceed; (*b*) promptness, flexibility, and fertility of suggestion; and (*c*) orderliness, consecutiveness, and appropriateness in what is suggested.[3]

[1] Irving Edgar Miller, *The Psychology of Thinking*, pages 148–150. The Macmillan Company, New York; 1917.

[2] John Dewey, *How We Think*, page 72. D. C. Heath & Co., Boston; 1910.

[3] *Op. cit.*, page 30.

Stated in terms of behavior, reflective thinking is dependent upon a stimulus in the form of a problem situation for which the individual has no ready reaction and which arouses a searching of experience for the right response. In order that this searching may result in a solution, there must exist:

(*a*) a rich field of associated experiences which the stimuli reinstate, with their attendant responses, one after another;

(*b*) activity of the stimuli in finding and selecting elements in the old situations that tend to meet the response demand and thus present a common factor linking the old stimuli with the new demand.

The result will be a conditioned response arising from compromises between old responses or responses aroused in new orders or combinations.

For illustration consider the problem, "What effect had the Emancipation Proclamation upon the outcome of the war?" A solution demands knowledge of various conditions existing both before and after its issuance:

(*a*) the repeated successes of Lee's army;
the continued existence of Southern forces in spite of the blockade;
the imminent possibility of recognition of Southern independence by England and France;
the opposition of the common folk of both these countries to slavery as an institution;
the effect of the blockade upon the English cotton mills;
the ambitions of Napoleon III.

(*b*) the weakening of the sustaining force back of the Southern army when the news of emancipation reached the slaves on the plantations;
the effect of the Proclamation on public opinion in England and France, making the struggle a moral

issue rather than one for sectional supremacy, and forcing the English political leaders to yield their desire for Southern recognition or for intercession.

To restate the case, learning is dependent upon building up experiences in rich fields of associated responses; these responses must be conditioned so that they will be, not reactions made in a random manner, but habits of discrimination that select the common or pertinent factor required for the solution of the problem. In the social studies the field of association includes basic information, and it must be conditioned to operate in such a way as to result in critical thinking.

Thinking must be conditioned in specific fields. The association of ideas upon which thinking is based must be built up in the specific field where the thinking is to function. A dairy farmer may have excellent power of critical judgment in regard to cows but be quite at a loss to decide whether the national government should subsidize the commercial air service. If he has studied national subsidies as they have operated in the past through the tariff and the "infant industries" of 1815, through the land grants to the railroads in the sixties and later, and through the independent growth of other commercial industries such as the telephone, telegraph, and oil companies, he will have a basis for deciding the relative values of direct and indirect subsidies, the responsibility of the government for encouraging commercial enterprises, and so on. But some knowledge in the related field is requisite if there is to be any selection or testing of evidence.

Responses should be orderly. Closely associated with the fertility of suggestions or associations in the immediate field in which thinking is to operate is another factor mentioned by Dewey — that of orderliness, consecutiveness, and appropriateness in what is suggested. It is a common occurrence

to have an idea flash into thought hours after it is needed, suggesting the exact point that would have clinched an argument or the retort that would have floored an adversary. Flexibility and orderliness of ideas are dependent on establishing systematic associations. The habit of marshaling responses to meet a given need, as one organizes a paragraph around a topical sentence, is a matter of training in habits of association. There is probably no teacher who has not met the answer to a question, "I know that, but I can't just tell you about it." The stimulus is there, the responses exist, but the requisite associations are lacking. It is significant to find the same emphasis on the need for definite information from one who has not been a teacher. The following comment is from an American woman who was the Berlin correspondent of the *Philadelphia Ledger:*

> I have been asked whether I did not encounter in Europe difficulties as a woman journalist concerned with so-called serious affairs. So far as I have observed, no single one of the many difficulties which have beset my path as an inexperienced adventurer in a new profession in strange lands arose from the fact of my sex. I encountered the difficulty of my own ignorance, of the fact that I had failed to absorb from my university training a vivid and accurate historical sense, a clear idea of economics, or fluency in any foreign language. I had been brought up on the easy-going theory that knowledge was "knowing where you can find a thing," and I had to learn that there is only one place where one can get information in an emergency and that is out of oneself. I had to begin to discipline my memory and rid myself as far as possible of a sentimental way of looking at things.[1]

Habits of systematic approach to history problems and the recognition of topical connections that subordinate the details to the main facts are valuable in establishing orderly associations. Haphazard attention and thinking that results in random associations do not tend to produce responses

[1] Dorothy Thompson, "On Women Correspondents and Other New Ideas," in *The Nation,* Vol. CXXVII, page 11; January 6, 1926.

that may be recalled at will. Memorizing definite facts with regard to specific situations and reviewing to test the ability to recall what was learned aid in establishing systematic responses.

Here is a group of dates that in themselves seem dull:

1807 The first steamboat sailed up the Hudson.
1811 The National Highway, or Cumberland Road, was begun.
1825 The Erie Canal was completed.
1829 The first steam locomotive was imported from England.

But bind these together as steps 1, 2, 3, and 4, indicating the surge to the country west of the Alleghenies of the population that had clung to the Atlantic coastal plain for two hundred years — a movement dependent on the development of new modes of transportation. Then compare the development of the past hundred years with the era 1620–1820; and new meaning has been given through this grouping not only to each event but to the associated whole.

Clear thinking requires organized associations. Clear thinking implies the establishment of proper associations between related responses. In the beginning of the learning process this association must be conscious; students must be taught to tabulate the causes and results of important happenings in order to develop this type of response. When the habit of seeking for such associations and recognizing them has been established as a response tendency to any new pattern of ideas, the mind ultimately tends to react to a new situation by a searching for connections with past experience and a resultant selection of the related ideas. This organization of ideas is a reaction that teaching should strive to make habitual in each particular field of learning. In English, for instance, a new word should meet with the habit-response of meaning, spelling, use, and — for advanced students — derivation. In history, ideas of social, industrial, or political

importance should meet the response tendency of relation — cause, result, connected events, location in chronological sequence, and connection with present and future problems of government. To establish such response tendencies, much experience in actual accounting for various political happenings is needed. Ideas must be specific, and there must be some basic information before relationships can be perceived. A child cannot see the relation of the Magna Charta to the Declaration of Independence if he has never heard of the former document. On this point a teacher of history speaks with some feeling:

> Let us decry "mere memory" as much as we like; the richer the associative memory, the more real the social or historical knowledge. What we decry is only meaningless or disassociated memorization; the whole value of the social sciences in daily living is that the present event calls up a complete network of memories. The mention of treasury refunding means to a well-trained student of society a mass of information on how our national debts were incurred, how previous national debts have been treated, the relations of our debts to European finance, and a vivid picture of the chain of action from taxpayer to bondholder; while to too many of our citizens it is a shadowy mystery.[1]

Youth does not object to learning. It is a mistake to assume that the acquisition of definite information is distasteful to youth. Students are eager to follow their interests, and if interest is combined with a school task that appears reasonably sure to bring success, effort is assured. If the normal child seems to be opposed to grappling with thought problems, it is largely because these problems are remote to him, because he has not the mental equipment to conceive the solution as possible, or because he lacks the store of experience or the study tools needed for attacking the new problem. It is the function of the school to relate

[1] Ruth Hardy, "New Types of Tests in Social Science," in *Historical Outlook*, Vol. XIV, page 327; November, 1923.

the materials of instruction to the student's experience and to provide exercises and training that will encourage reflection. Differences in ability to think abstractly are due not only to variations in native equipment but also to differences in experience — both vicarious and actual experience. Some students can get a keen sense of reality from an imaginary participation in past events, and it is the work of the history teacher to see that the imagination is supplied with details that make the past vivid and reinforce the connection of the past with the present through the medium of the youth's own experiences. A certain amount of information is essential for this undertaking. The student possessing the necessary ability does not object to hard study if he finds purpose and satisfaction in it.

> The zest in doing a task is the highest attraction to children in a normal environment. Above everything else they delight in doing something that is serious, just as they see adults working at their daily tasks. They can develop an interest in doing a job as a real job should be done.[1]

The teacher who has seen the amazing energy with which young people attack a project that they feel to be their own, such as reproducing the Constitutional Convention and impersonating the leaders, or collecting pictures and data relating to local history, recognizes the existence of a tremendous store of energy that is only occasionally tapped by the demands of the school. The problem of the history teacher — and it is a task demanding complex skills — includes aiding the student to find satisfaction in the task of gaining basic information.

Facts considered as a means, not as an end. No educator will claim that the memorization of isolated facts is learning. Facts in themselves have little value or significance; to be

[1] W. H. Burnham, *The Normal Mind*, page 209. D. Appleton & Co., New York; 1924.

meaningful, they must become part of the individual experience and be related to countless other facts. Skillful teaching is necessary if these relations are to hold permanent value; but it is just as true that connections will not carry deep significance unless some facts are known. The following comment on the relation of factual knowledge to the problem method in teaching is pertinent:

> Of late some objections have been raised to the theory that the problem is the starting point of work on a portion of geography. Mr. Stull, associate in geography at Teachers College, Columbia University, has stated in class that children at the beginning of their study of a country do not have a sufficiently wide factual background to enable them to select problems intelligently. In discussing the same subject recently, both Professors McMurry and Hillegas have made the same sort of statements, giving as their opinion that it is likely that a problem of real value can be discovered and stated only after a careful preliminary study of a subject.
>
> Although there are no data to prove it, it stands to reason that if the pupils had had the common body of important and essential information which would have resulted from an early study of the text, they would have had a better basis for evaluating and criticizing the reports that were given than they show evidence of having in the lesson under discussion; and those who reported would have had a better foundation for weighing the worth of the material on which they were ready to report.[1]

This need for information with which to attack a problem has not always been stated by the "progressive" school. Yet here is the testimony of Dewey, who has done more than any other educator of today to emphasize the fact that learning functions only as it becomes part of the child's experience. In continuing the paragraph quoted previously, in which he asserts that "thinking cannot go on in a vacuum," he adds:

[1] Edwin H. Reeder, *A Method of Directing Children's Study of Geography* (Contributions to Education, No. 193), page 12. Teachers College, Columbia University, New York; 1925.

But there is all the difference in the world whether the acquisition of information is treated as an end in itself, or is made an integral portion of the training of thought.[1]

The linking of attitude, information, and reflection. To make effective an attitude of interest in civic affairs there must be some knowledge about the structure of government. Interest is sharpened by understanding. Throughout the course of American history there are incidents which may be used to link an attitude of civic responsibility with historical information interpreted in relation to the student's own experiences. As an illustration of this, consider the Hayes-Tilden controversy, which, while of little importance today as far as the leading characters are concerned, is significant in connection with the working of the electoral college. To understand the situation the class should possess definite information as to the way in which the President is elected, the difference between the electoral and popular votes, the variation of registration regulations in different states, and so on.

This information can be applied in the student's own experience with regard to the registration of parents, to the issues in a present-day campaign, and to whatever else in the immediate political situation will give reality to the student's feeling about voting. If the immediate campaign is local rather than national, the approach may be through the city or state requirements for registration; a search for the derivation of authority from the Federation will lead to a study of the Constitutional regulations. The emotional tone that does much to make interest and attitude lasting may sometimes be furthered by the feeling of success that comes to the child who persuades an indifferent adult to register, or through the activity that accompanies preparation for a

[1] John Dewey, *How We Think*, page 52. D. C. Heath & Co., Boston; 1910.

school election and participation in it on the real "election day." It is useless and wasteful to attempt to build up proper attitudes toward the ballot without seeing that the students gain the information necessary to make such attitudes practical in related situations of adult life.

Training is needed for voters. It cannot be claimed that skill in government or a knowledge of its structure and ideals is a product that springs spontaneously from residence on American soil. In its infancy our government was organized and maintained by means of a great emotional need that aroused united action. Policies were determined by a highly selected group who represented the conservative element in the population, where voting qualifications were so rigid as to restrict suffrage to a fraction of the adult male population. With those restrictions gone, with the emotional stress relieved, and with the presence of a greatly increased voting population made up of varied inheritances, there is grave need for more serious consideration of the training of voters. Munro, an accepted authority in the field of political science, has expressed this need as follows:

> A democracy gives the ballot to every adult citizen, yet the ballot is potentially the most dangerous weapon that can be put into unthinking hands. As a weapon of vindictiveness, oppression, and nation-wide folly there is, potentially, none other to equal it. That is why it has been said, and with grim truth, that we are today engaged in a race between education and disaster, between the rule of reason and the misrule of emotionalism.[1]

Unthinking emotionalism sways the mob mind. Mob action is possible only when critical reason fails to control action. Unless habits of deliberation are properly developed, there will be little resistance to the mad emotionalism that the student may encounter in adult life. If habits of critical

[1] William Bennett Munro, "Scientific Education and Unscientific Democracy," in *Harvard Graduates' Magazine*, Vol. XXXIV, pages 175–185; December, 1925.

thinking are not built up in school, they will not exist for the mass of voters. False enthusiasm for impractical Utopias, and the advocacy of measures that make for immediate personal gain find ready acceptance in the mind of the uncritical citizen. Shays's " army," as well as the later one led by Coxey, was made up of men clamoring emotionally for money, not men trained to study causes when seeking remedies for discontent. On the other hand the Constitutional Convention saw a gathering of deliberate thinkers who recognized an emotional menace as well as political inefficiency in a situation that they sought to correct through constructive legislative procedure. As long as industrial prosperity continues, unthinking voters will follow the party that has the most powerful machine and the one whose promises sound most alluring. But if our nation were to become involved in an economic disaster such as recent years have brought to certain formerly stable countries in Europe, an accounting would be called for. Whether emotionalism or deliberation will control the electorate depends in large part on the kind of training for citizenship given in the schools. Judgments based upon unsupported opinion give way too readily to emotionalism. It will require training in habitual responses to deliberation to offset the lowered threshold of emotional appeal. Deliberation based on a demand for evidence can take place only when there is some basis of fact for thinking — some appreciation of accuracy and proof. A scientific attitude of mind, therefore, in matters of government must come through careful training in the social studies, and it must be a training based on the acquisition of some definite information as to problems of the political past and present, and on the exercise of reflective thinking about these problems in relation to the evidence in the case.

Summary. Improved behavior in social and civic situations is the ultimate aim in teaching history. The attitudes

and ideals that determine behavior are the result of ordinary experience and of direct training. To be socially effective, training that is intended to develop adaptable ideals and attitudes must link school experience with situations of the type that the student will need to face in later life. Emphasis on attitudes is frequently associated with certain fallacious deductions — that information is useless and that the attainment of right attitudes supplants any other type of learning. Attitudes are built by experiences, and the most effective civic experiences are those that involve the assumption of responsibility in a school democracy. Effective citizenship demands the correlation of deliberative thinking with such experience.

Neither analytical nor constructive thinking can take place in isolation from the acquisition of knowledge. One cannot think in civic situations without having some information about the facts of government. Reflective thinking is dependent upon a problem situation; and the successful search for the solution of the problem is in great measure dependent upon the existence of a rich field of experiences related to the problem. Thinking is most effective when it is specialized through practice in specific fields.

The learning of historical facts is valuable in so far as these facts are related to individual thinking in regard to social and political issues. Knowledge of the past needs to be linked with present problems through the student's own experiences, for the habit of meeting civic issues with deliberative thought is most important in a democracy.

CHAPTER THREE

Characteristics of New-Type Tests Valuable in the Social Studies

The new-type tests present two immediate differences from the older form: The mechanics of answering are reduced to a minimum by the system of checking the correct response among the several suggested answers; there is included a much larger sampling of questions — from ten to twenty times as many questions as are possible with the essay-response type of test. The advantages of the new test forms have brought a reconsideration of the use of examinations.

The purpose of examinations. A few years ago there was a decided movement in opposition to examinations of any type. This attitude was a natural reaction from the abuses consequent upon making promotion depend upon a single final examination, and was a practical protest against the situation that college entrance examinations forced upon the lower schools. Since that time there have been experiments, not altogether satisfactory, in basing promotion solely upon the teacher's recommendation. There is no occasion here for a discussion of that issue; it is sufficient to say that the experiments have been of value and that the teaching body is now in a better position to evaluate the reasons for using examinations.

It is not without significance that the president of a university with the scholastic tradition of Harvard should write on "The Art of Examination." In his article Lowell thus characterizes the purposes of examinations:

> Certainly we have not considered with sufficient care the object, scope, and utility of examinations. May they not be said to have three distinct objects: (1) To measure the progress of the pupils; (2) as a direct means of education; (3) to set a standard of achieve-

ment? There have been examinations that required merely committing to memory masses of useless data, and made drudges of those preparing for them; but that does not mean that a potent force for good should be discarded because, if not intelligently used, it is capable of abuse. It would be like arguing that, because there have been bad laws, legislation should be abolished.

The conclusion to be drawn from the views here presented is that examinations properly used are a vital part of the educational process, but that the art of using them to produce the best results is highly complex and difficult. . . . To make a good examination paper is far more difficult than is commonly supposed. To do so requires much time and thought; but upon no part of the educational process can time and thought be better spent.[1]

To measure, to educate, to set standards — thus are summarized the varying and yet unified purposes of examinations. It is quite as significant in the use of new-type tests as of older forms to recognize that measurement alone does not comprise the purpose of examination. In fact, the purposes set forth above may be combined in the second and dominant one — testing to aid in instruction. Concerning the use of the essay-type examination, nothing finer has been written than Johnson's discussion of the open-book test.[2] There is shown the value of the examination that summarizes the steps already taken, requires additional learning, and measures power as well as knowledge. The usefulness of such an examination to the teacher is limited only by the difficulties attendant upon the essay response. But these difficulties are serious enough, including for the student the marshaling of evidence as well as the correct expression of ideas, and for the teacher the time and energy consumed in reading papers and the unsatisfactory nature of the subjective estimate. Nevertheless, the essay test has a place in the social studies and will profit as the present emphasis

[1] A. Lawrence Lowell, "The Art of Examination," in *Atlantic Monthly*, Vol. CXXXVII, pages 58–66, January, 1926.

[2] Henry Johnson, *The Teaching of History*. The Macmillan Company, New York; 1917.

upon objectivity is reflected in improved expository responses.

Differences in the essay-response and new-type examination. The functions of the essay-response and new-type examination in the social studies should be distinctive. The essay form is valuable for the expression of individual opinion and interpretation, for the discussion of controversial issues, and for summaries to show comprehension of the subject as a whole and its interrelations with other topics. But as a basis for forming comparative estimates of student learning the scores derived from an essay examination are too highly subjective to be a definite criterion. This is not saying that the objectivity of the essay test cannot be increased nor that the objective test cannot be used for the measurement of opinion. But economy of the history teacher's time indicates that, until both forms are further developed, the one is preferable when subjective responses are desired, the other when objective responses are called for.

To make the essay response more definite requires provision for exercises in which the student summarizes his answer in a single, concise statement or with brief sentences, each of which develops a specific point relating to the topic. In addition to the requisite knowledge, successful answers of this type demand not only clear thinking but skill in handling language — skill in packing meaning into a few words. Quite aside from the time needed to train students in correct expository responses is the drain upon the teacher's time and energy in marking the papers. And even when completed, such marks are highly subjective and variable; while to establish definite criteria for scoring increases the time used. To make expository responses more definite, not only must questions be planned in relation to certain answers that the teacher hopes to secure but, after the test is given, responses must be evaluated on the basis of pupil learning

instead of teacher expectation by assembling the best and the poorest responses and determining gradations of success. There must be careful discrimination between different elements of the response, such as content, neatness, grammatical structure, and paragraph unity, while percentage values must be assigned for each item. There can be no question of the values resulting from this training; but the development of this skill is the particular responsibility of the class in English composition. Adequate teacher time should be allowed for it there, since the history teacher needs to emphasize other and different aims. The desired objectivity in examinations may be secured more readily if the teacher turns directly to a form designed to obtain that quality.

Assuming that the teacher wishes to use the results of examinations to further the learning process, it is pertinent to ask what further significance objective tests in the social studies have. And the answer is: Comprehensiveness, reliability, and validity resulting from scientific construction have brought new uses to testing, as well as greater significance to the interpretation of test results.

Comprehensiveness. The distinctive advantage of the objective test is in its comprehensiveness, for it makes possible responding to fifty or one hundred questions in about thirty minutes. It should be remembered that any examination is merely a sampling of the field of knowledge that the test attempts to measure. Hence the student's chance for finding in the range of questions those that represent samples of *his* knowledge increases with the number of items offered. Likewise the teacher's chance for obtaining a fair estimate of what the student knows and does not know increases with the number of items offered. To some extent this greater range is due to the time saved for the student by the mechanical efficiency implied in indicating answers by underlining

one of several phrases, checking a significant numeral, or supplying a key word to complete a sentence in place of writing sentences or paragraphs.

Brinkley's study at the Horace Mann High School in New York City compared the comprehensiveness of the two types of examinations. Twelve minutes was allowed for a true-false test of fifty items on the Federal Constitution. Two weeks later the same class was given a fifteen-minute essay test in which each pupil was to select one of the following topics:

The compromises of the Constitution.
How a bill becomes a law.
How a federal nation was formed from a confederation of states.

An analysis of information items in the responses showed the following results:

For the true-false test the average number of items responded to was 49; for the essay test the average was 12, while the smallest number given was 5, and the largest number was 22.[1]

In another investigation that compared the amount of information in a given essay response with that in new-type questions, two different tests on the same topic were given to fifty-four students. In commenting on the experiment, the investigator says:

In the essay examination the students knew half as much as on the other when checked off against the points that they had been given in the course of the subject under investigation. . . . The average student knows twice as much of the subject when tested by an objective information test as when tested by the essay type of examination.[2]

[1] Sterling Brinkley, *Values of the New-Type Examination in the High School, with Special Reference to History* (Contributions to Education, No. 161). Teachers College, Columbia University, New York; 1925.

[2] Donald A. Laird, "A Comparison of the Essay and the Objective Type of Examination," in *Journal of Educational Psychology*, Vol. XIV, pages 123–124; February, 1923.

The scope of the essay examination as a measure of the student's knowledge of historical information is restricted by the difficulty of expression, and the student is unable to do justice to himself.

Comprehensiveness increases the range of distribution. There is a further significance in this comprehensiveness, due to the fact that a fifty or one-hundred point test makes possible a scaling of attainment by which the bright student as well as the dull one may be measured fairly accurately. Differences in achievement may be determined both by the number of correct responses and by the degrees of difficulty represented in the questions. In large classes this point is of much advantage to the teacher in estimating individual difficulties in learning, as well as in providing an objective basis for ranking members of the class.

Comprehensiveness increases reliability. For a test to be meaningful it should be reliable; that is, the error in the score due to varying factors should be very small. Statistical studies show that reliability increases as the number of questions increases and in about the same ratio. In tests, the reliability indicates the chances for obtaining a similar score if the pupil were to take the same examination again under the same conditions. In other words, it indicates the extent to which a pupil's score is in error. To determine reliability, a long examination is frequently broken in halves and the scores for the odd-numbered questions are compared with the scores for the even-numbered questions; or the whole examination is given again to the same students within a comparatively short time; or two equivalent forms of the entire test may be used.

The degree of reliability is expressed by a coefficient of correlation. For this measure, 1.00 indicates perfect correspondence between two sets of scores, while varying degrees of correspondence are expressed in decimal fractions, both

positive and negative. A complete lack of correspondence is indicated by .00, while an absolute inverse correspondence, an entirely negative correlation, is expressed by − 1.00.[1]

In brief, statisticians are trying to find a measure of dependability in examinations. Whatever tends to make the responses less variable contributes to greater reliability in tests. Factors that increase reliability are (1) greater length, both as to number of questions and time required for responding, (2) range of difficulty, provided the items are not unreasonable, (3) objectivity of responses, (4) uniformity and accuracy in scoring, and (5) arrangement of the test in several units, each having a time limit. Through its adaptability to these conditions the new-type test has proved its reliability.

Evidence of reliability. The investigation conducted by the Modern Language Association provided data to show the reliability and the validity of the new-type test.

> The general conclusions from the data of this experiment are that the new-type examinations are roughly twice as reliable and valid as the old-type examinations of equal time allowance; that the new-type examinations afford comparable measures for all classes in a given subject in the same and different years and thus offer a means of eliminating overlapping of classes and varieties in local school standards to a much greater extent than they are eliminated by the old-type Regents Examinations[2]; and that the new-type examinations over a series of years will cost not more than 10 per cent as much as the old-type examinations, as administered and read by the College Entrance Board, cost.[3]

[1] However, degrees of correlation, or correspondence, are not proportionate to the correlation coefficient. Thus a correlation of .50 indicates much less than half as close correspondence as a correlation of 1.00.

[2] These are examinations issued by the New York State Department of Education by authority of the Board of Regents.

[3] Ben D. Wood, *New York Experiments with New-Type Modern Language Tests* (Publications of the American and Canadian Committees of Modern Languages, Vol. I), pages 318–319. The Macmillan Company, New York; 1927.

The unreliability of the old-type examination in the social studies is discussed at length in the study conducted by graduate students of the University of Iowa under the direction of Ruch.[1] This study, which was financed by the Commonwealth Fund, is one of much importance in the development of better examination methods, and the whole of this brief and easily read volume is recommended to teachers of the social studies. In summarizing the investigation of the reliability of state eighth-grade examinations of the old-type, the following startling conclusions are stated:

(1) Sixteen state eighth-grade diploma examinations representing eleven different states yielded average reliability coefficients as follows:
   (*a*) 0.62 when the examination was constant and the scorer (reader) was the variable.
   (*b*) 0.43 when the scorer was constant and the examination was the variable (1923 versus 1924 examinations).
   (*c*) 0.38 when both the scorer and the examinations were variables.
(2) The agreements represented by correlation coefficients of from .38 to .62 are approximately $7\frac{1}{2}$ per cent to 20 per cent better than chance assignment of marks.
(3) The unreliability of examinations is due to two general causes; viz., limited sampling and subjectivity of scoring.
(4) Unreliability due to subjectivity is completely or almost completely eliminable.
(5) Unreliability due to limited sampling may be minimized by the use of examinations composed of 100 to 250 or more short questions rather than five to ten broad questions; i.e., by adopting an *extensive* plan of sampling rather than an *intensive* sample.[2]

Conclusions 6 to 10 are not quoted.

Objective tests are more valid. The validity of a test depends upon how well it measures the ability it is designed

[1] G. M. Ruch and Others, *Objective Examination Methods in the Social Studies.* Scott, Foresman & Co., Chicago; 1926.
[2] *Op. cit.*, page 21.

to measure. It is, of course, evident that the new-form examination much more nearly fulfills this purpose in the field of history, civics, and geography than does the essay type. The fact that a measure of historical information is provided as distinct from ability to express oneself assures the teacher that the time spent in answering such a test is devoted to the recall and interpretation of information in that particular field. To the student it means that his whole effort may be concentrated on the recall of knowledge in history rather than divided between this endeavor and the struggle to express his thought with ease and conviction. Undoubtedly there are times when the latter ability should be tested, but it should not be forgotten that in such cases it is upon his skill in composition as well as upon his knowledge of history that the pupil is being examined. Ability in written expression is a complex process involving many other skills more intricate than mere sentence structure. Not the least of these is the selection of the salient points out of that vast chaos of "all he knows," and the organization of the selected facts with due recognition of the relation of primary and subordinate points to the main topic. Because of its complexity this skill is one in which the youth is slow to develop mastery; even more, this skill of necessity requires preliminary information. Testing in history should deal with the acquisition of this information and with its interpretation and significance rather than with skill in expression. The validity of an objective test, which permits concentration on the recall of subject matter, gives rise to the students' comments that these tests are easier and induce less fatigue.

While the objective test possesses increased validity because it measures knowledge of history as distinguished from knowledge and skill in the use of language, validity in a history test implies much more than the measurement of historical information. How much more is immediately evident from

any statement of subject aims. Without doubt the measurement of the "higher objectives" is more difficult than measuring factual information. The difficulty is not one that is characteristic only of the new-type test, however; it is a difficulty, as noted in earlier discussion, inherent in the subject matter itself. Better types of tests will be more readily developed by the experimentation of the actual teachers who combine an understanding of the goals of history teaching with skill in handling the technique of the new-type test. The results of the experiment in Rochester [1] give ample proof of this in their sample tests, showing a significant advance in applying the new question forms to the social studies test.

When instructors realize that objective examinations bring comprehensiveness, reliability, and validity to the measurement of achievement, improved methods of instruction will result as new values for tests are found in measuring learning difficulties, in making instruction more personal, and in motivating students toward higher standards.

Summary. Examinations are being reëvaluated in the light of certain marked advantages that result from the objectivity of the new-type test. Purposes are conceived in broader terms than merely to measure achievement; the primary purpose of an examination is to serve as a direct means of instruction as well as to measure progress and to raise standards. In the social studies the functions of the essay-response and the new-type test may well be distinctive. The essay form is valuable for the discussion of controversial issues and individual opinions, and it will become more valuable as its objectivity is developed. However, such development requires more time than the teacher of history should take from her special responsibilities. The new-type

[1] Alice N. Gibbons, *Tests in the Social Studies: A Record of a Testing Experience in Senior High School Social Studies.* The National Council for the Social Studies, Athens Press, Iowa City; 1929.

test presents in forms readily adaptable for social science tests a medium for objective measurement. This objectivity results in a high degree of comprehensiveness, reliability, and validity in examinations. As the significance of these qualities in examinations is realized, new uses for testing will bring improved instruction.

CHAPTER FOUR

Using Tests to Improve Instruction

In previous chapters testing was considered as an aid to the progressive teacher who would make sure that her students are gaining a basis of historical fact for their problem study; factual knowledge was analyzed in relation to reflective thinking; and the purpose of examinations was considered in relation to the advantages resulting from the objectivity of the new-type tests. Since the dominant purpose of examinations is the improvement of instruction, the achievement of this aim will be considered through the use of tests for (1) the diagnosis of learning difficulties, (2) individualizing instruction, (3) raising students' standards, and (4) providing a better measure of achievement.

THE DIAGNOSIS OF LEARNING DIFFICULTIES

The use of tests for the diagnosis of difficulties in learning history has not been commercially developed, although this is a promising field for investigation. However, excellent standardized tests for that most essential skill, reading comprehension, are available.[1] Reading tests are easily administered even by a teacher with no previous training in testing; while the interpretation of scores is simplified by norms established in terms of reading age or grade, from the elementary school through high school and college. At the beginning of the term it is especially helpful to a teacher to know what ability or deficiency each student has in read-

[1] Haggerty Reading Examination, Sigma 3, for Grades 6–12. World Book Company, Yonkers-on-Hudson, New York.

New Stanford Reading Test, for Grades 2–9. World Book Company, Yonkers-on-Hudson, New York.

Thorndike-McCall Reading Scale, for Grades 3–12. Teachers College, Columbia University, New York.

Whipple High School and College Reading Test. Public School Publishing Company, Bloomington, Illinois.

ing; for it is absurd to expect the same degree of textbook comprehension or collateral reading from students whose reading abilities vary from that of the seventh grade to that of the twelfth. Such variations of ability occur more frequently in the high school of today than in the past, because the compulsory attendance age has been extended. An ideal situation would be to place those students of limited reading ability in a special section where assignments might be adjusted to their skill in reading and to the development of increased power.

Since reading comprehension is so largely a matter of training and application, it is an advantage to the student to know his rating. While students of inferior mental ability will always be poor readers, much can be done by training to increase both the rate of reading and comprehension for the average student. An attitude of indifference to reading, or of discouragement about learning, may thus be changed to an active effort for improvement through the recognition of a difficulty that is remediable. Students show much interest in having their vocabularies measured; and since the scores show high correlation with reading tests, the teacher, too, gains much from this measure. There are available carefully constructed vocabulary tests [1] giving norms for both high school and college achievement. The cost of such a test is so slight that even if the teacher herself must pay for it, the results will compensate for the expenditure. Two excellent reports of special vocabularies in the social studies field are on record.[2]

[1] The Inglis Vocabulary Test (designed for high school and college). Ginn & Co., Boston.

Thorndike Test of Word Knowledge. Teachers College, Columbia University, New York.

[2] Mary S. Gold, "Testing Vocabulary in History," in *Historical Outlook*, Vol. XVII, pages 285–291; October, 1926.

Orlando W. Stephenson, "The Special Vocabulary of Civics," in *Journal of Educational Research*, Vol. XVIII, pages 297–304; November, 1928.

If a standard test cannot be secured, an estimate of comparative reading ability in a class may be obtained by using carefully prepared questions upon a paragraph of expository matter. If an objective answering scale has been arranged, then listing the scores in order of size may give a fairly satisfactory ranking of reading ability within the class. It is an advantage to make such an informal test brief so that it will not consume more than the minimum of pupil time, perhaps about twenty minutes. At the beginning of the term students cannot be expected to understand the significance of too much diagnostic testing, and the instructor who desires to measure different skills will need to avoid using too many tests or tests that are long enough to tire the student before he has attained a compensating interest. From her own standpoint, too, the teacher will need to economize time. The new term presents so many demands for determining units and outlining work plans that the clerical work of test correction should be reduced to a minimum.

The Rochester study [1] includes a copy of a reading test that was most carefully devised by social science teachers. For diagnostic purposes the test seems rather long, but judging from the similarity of approach and of norms in Parts I and II, it would seem possible to shorten it by omitting either one of these sections. It is significant to note that of the 1961 high school students who took the test, the scores showed a variability of reading comprehension ranging from 9 per cent to 100 per cent, with a median of 66.5 per cent.

As an illustration of further value that the social science teacher may gain from the reading comprehension test, the following comment made by Miss Gibbons is cited:

> The low rank of tenth grade students in comparison with the eleventh and twelfth grades, as measured by this test, tends to show

[1] Alice N. Gibbons, *Tests in the Social Studies: A Record of a Testing Experience in Senior High School Social Studies*, page 78. The National Council for the Social Studies, Athens Press, Iowa City; 1929.

that reading ability is decidedly better in the upper grades. Therefore, it would seem unwise to continue the present custom of allowing tenth grade students to be placed indiscriminately in the same classes with eleventh and twelfth grade students. The two and one-half per cent difference between eleventh and twelfth grade students seems not enough to warrant class separation of these two grades in social science work. The low rank of tenth grade students in comparison with those of the eleventh and twelfth grades may be in good part due to the fact that there is a much more selected group in the twelfth grade. In comparing types of classes, statistics show that the tenth grade (average) classes in the technical arts department attained lowest rank. This tends to confirm the subjective opinion held by some teachers that the unusual number of failures of technical students in social science may not be primarily due to a lower grade of mental ability, but to lack of reading ability, having its origin in indifference or scorn of books as tools of learning.[1]

The most valuable diagnostic test for the social studies is the Working Skills Test,[2] compiled in the course of this same experiment in Rochester. The test was published in its original form as a seven-page booklet having five exercises with subdivisions and items that total 100 points. In the directions to the students the statement is made:

> The purpose of this test is to learn something about the skills and abilities in handling materials you are bringing to the social sciences.

The titles of the five exercises are:

I. Becoming Informed about a New Book
II. Use of General Reference Books
III. The Reading of Charts, Tables, and Diagrams
IV. Newspaper Reading
V. Summarizing and Outlining

Exercise II is especially suggestive in its attempt to estimate the student's ability to use reference aids:

[1] *Op. cit.*, page 40.

[2] Working Skills Test, constructed by E. Riley of East High School, Rochester, New York. Awarded first prize in a recent contest, and published in Ruch and Rice's *Specimen Objective Examinations*, by Scott, Foresman & Co., Chicago; 1929.

Use of General Reference Books

The degree to which a library is useful to you as a student is determined in part by the ease with which you can obtain the facts and information you want. Below are two lists. One is a list of information you might want to find. The other is a list of general reference books. Place the letter of the reference book (i.e., A, B, C, etc.) in which you find the information in front of the number of the information asked for.

LIST OF INFORMATION DESIRED

____ 1. To find the source of the Volga River.
____ 2. To look up recent articles about Mussolini.
____ 3. To find the value of the exports and imports of the Port of New York for 1925.
____ 4. To obtain a good description of the western expansion movement in the United States.
____ 5. To find the number of students in American universities.
____ 6. To find out if there is a railroad in Moscow and Warsaw.
____ 7. To find out what has recently been written in magazines about evolution.
____ 8. To find a complete, technical, comparative discussion of the difference between a partnership and a corporation.
____ 9. To locate the Ægean Sea.
____ 10. To look up brief information on Gothic architecture.
____ 11. To obtain a careful discussion of the rights and obligations of American citizenship.
____ 12. To find the total wheat exports from the United States in the last ten years.
____ 13. To look up the main facts of Shakespeare's life.
____ 14. To find the difference in meaning between two words.

Other items of information numbered to 20 follow, after which is given:

LIST OF REFERENCE BOOKS

A. Atlas
B. *World Almanac*
C. Encyclopedia
D. *Who's Who in America*
E. *Readers' Guide to Periodical Literature*
F. An American history book
G. A civics book
H. English language dictionary
I. An economics book

It is evident that a test of this type educates while it measures, and that its value for diagnostic purposes at the beginning of the term does not limit its use a second time at the close of the term as a measure of progress. Such a test can be used in any year of junior or senior high school and is as applicable to one locality as to another. It is to be hoped that the test will be revised in line with the modifications outlined by Miss Gibbons and made commercially available. Such experimentation as resulted in the Working Skills Test points the way to a development in social science testing that is most worth while. A helpful suggestion for this type of test is to be found in Briggs's Dictionary Test.[1]

The diagnostic tests so far published for the social sciences have not proved satisfactory for one reason or another. The diagnosis that will aid the instructor meeting a new group of students is not so much a measure of their knowledge of history as of their methods of approaching the study. The Barr Diagnostic Tests in American History [2] are a suggestive compilation, but of little use to the teacher until the close of the term as the tests are based upon a knowledge of history. The Barr tests are in five sections, covering "comprehension, chronological judgment, evaluation of facts, and causal relationship." The responses are not entirely objective, and the four-page pamphlet requires ten minutes or more to correct. The Kepner Background Test,[3] while sometimes called a diagnostic test, is again a measure of the information with which a student approaches the social studies. The test covers the field of social studies in general and is entirely objective, requiring only a short time to score. Such a survey is valuable if the instructor has the opportunity to

[1] Thomas H. Briggs, "A Dictionary Test," in *Teachers College Record*, Vol. XXIV, pages 355–365; September, 1923.

[2] A. S. Barr, Diagnostic Tests in American History. Public School Publishing Company, Bloomington, Illinois.

[3] Tyler Kepner, Background Test in Social Science. Ginn & Co., Boston.

group students into sections according to their previous knowledge of the subject, and it would be valuable for classifying students who enter high school without credentials. If measurement is desired for classification in either American or European history, the tests of the Columbia Research Bureau are the best available.[1]

Preliminary tests given at the beginning of the term have an added value if they are so constructed that the teacher may use the same or a similar form at the close of the term to measure her own teaching accomplishment. But a pre-test that is based on historical information only is likely to discourage the student without giving the instructor any special teaching guidance. However, if the test is based on skills or powers which the student can see as worth-while goals of study and which may be measured at a later time to show advancement, the pre-test is amply justified. Miss Gibbons, after commenting on the amount of work entailed in the statistical tabulation of background tests, summarizes the teachers' attitude thus:

> The writer has noticed that a doubt of the value of factual pre-testing seems almost universal among teachers. If they do become convinced of the value of this type of test for immediate classroom use, it is invariably through the use of very short pre-tests on the topic or unit being studied at the time. General pre-tests covering the subject matter of a year, a term, or even a two-month period are mainly considered worthless.[2]

[1] Columbia Research Bureau American History Test. World Book Company, Yonkers-on-Hudson, New York.
American Council European History Test. World Book Company, Yonkers-on-Hudson, New York.
Junior American History Test. World Book Company, Yonkers-on-Hudson, New York.

[2] *Tests in the Social Studies*, page 39.

TESTING TO INDIVIDUALIZE INSTRUCTION

To see the students as individuals is an essential goal of modern teaching. Given a measure of a student's ability in terms of the established reading norms, and of his knowledge of study tools, such as reference books, the instructor will have a very good substitute for the complex estimate expressed by the IQ. Perhaps she will have a more usable estimate, for the intelligence quotient must be treated with caution and interpreted skillfully. The intelligence quotient derived from scores obtained from a single group test is a significant indication but by no means an absolute measure of intelligence. It is seldom that a school system uses a sufficient variety of group tests to strike a dependable average score, and more rarely that a public school can afford to measure students by a revision of the Binet-Simon Scale for Measuring Intelligence, which must be given individually by a trained expert. Furthermore there is the controversy over the point of mental maturity — the age which marks the full development of power to learn, as distinguished from achievement in learning. This question, which is still debated by psychologists, makes the use of intelligence tests much less satisfactory at the high school age than in the elementary school. Then, too, many high school principals are chary about passing on to teachers records of students' IQ's. And though at first it seems contradictory to principles already asserted to state that tests should be given without making the results available to the instructors, it is wise in this case, considering the dangerous ease of thus hastily or ignorantly classifying students with a measure requiring careful interpretation.

Granting, then, the limitations of the group-test IQ, the teacher who approaches the student with other diagnostic measures, such as those indicated above, possesses a

usable estimate of his preparation for the course. The estimate is usable in the double sense of having significance not only for the teacher but also for the student who recognizes that he has deficiencies which are in need of correction. Seeing the students as individuals thus leads to definite steps in teaching them as individuals.

Improving individual study habits. The most valuable contribution made by the several plans for individualizing instruction is found in the development of better study habits. Improved preparation is secured through the use of clear and full assignments given to the student in mimeographed pages or printed work sheets, and through the frequent use of self-measuring tests. Even though it is not desired to adopt the individualized instruction plan in its entirety, the best features should not be ignored by teachers. The system of keeping on file tests that deal with particular units of work and using these for individuals or groups of students who have completed a certain assignment is at once an incentive to student achievement and a time-saving device for the teacher. If one or more alternate forms are devised when the original test is constructed, the different forms may be used repeatedly with little danger that the test content will be passed from one student to another. This statement, however, should not be applied to examinations that are a weighty factor in promotion. The latter point and the adjustment of tests to the " term contract " will be considered in the discussion of test construction. To build habits of good study, the frequent and definite questioning about preparation, which objective testing makes possible even for the busy teacher, is most essential and valuable.

Tests individualize teaching. To plan lessons so that study will mean the establishment of systematic and regular habits of application and to arrange for a reliable and frequent checking of all students places a heavy tax on the skill and

ingenuity of the teacher. For it is an unfortunate paradox that the necessity for using tests to check students' work increases with the number of students and the size of the classes for which a teacher is responsible. With the natural competition from other studies and from extracurricular activities, the students in large classes often neglect the lesson preparation of those courses in which a regular accounting is not required. Under these circumstances some such follow-up as the written quiz is almost a necessity. A brief experiment [1] conducted to study this problem showed that when a written question was given a geography class each day the class average on a ten-minute test was raised from 41.3 per cent to 78.3 per cent after four days of repeated testing. More can be ascertained about lesson preparation in a carefully prepared ten-minute test of the brief-answer or true-false type than in an hour of repetition of textbook matter.

To the busy teacher who frequently has as many as two hundred students reporting each day, the saving in time made possible by the ease of checking objective answers is an item of much significance. The situation is further improved for the teacher by the fact that objective responses can be corrected by a clerk or student assistant or by the pupils themselves in the class period. It is true that some of the time saved in correcting papers is at first consumed in more careful formulation of objective questions; and such questions must be carefully planned — not only to use the new question forms with discrimination but to select points of significance and eliminate petty items. But it is certainly in line with better method to spend the time on the preparation of an examination rather than on its correction, and, of course, experience reduces considerably the time required to construct questions. Since tests may be used term after

[1] Mendel E. Branom, *The Measurement of Achievement in Geography*, page 10. The Macmillan Company, New York; 1925.

term if pupils are not allowed to retain copies and if some provision is made for variation, the burden of preparation is slight after the first copies have been arranged. Improved examinations result from the successive winnowings to which the questions in old tests are subjected by repeated use. Coöperation of teachers in the construction of tests saves the individual's time and improves the result.

Checking individual reading records. The objective test is a convenient means for checking the completion of the collateral assignment. The regular checking of notebooks is commonly neglected, not only because teachers have too much else to do but also because of the monotony of the task. It is wiser to require nothing but original work in the student's notebook and to limit such assignments to the amount that can be immediately and carefully checked. While testing may not be so valuable to the student in this connection as taking notes, the time and energy saved to the instructor and the elimination of the student's temptation to copy notes more than counterbalance the loss. Weekly or monthly tests on collateral assignments may be valuable, too, in emphasizing the fact that the purpose of reading is not accomplished by the mere taking of notes. The teacher should remember, however, with special emphasis in this connection, the danger of including too many questions of small detail in brief-answer tests. The purpose of collateral reading should be to stimulate interest and give a better understanding of past life; there should be no emphasis on memorization of details. Hence the questions on such reading should be general in nature and fairly easy; the best type of questions are those that emphasize the interpretative value of material rather than its reproduction. If the questions are wisely chosen, they should lead students to take a more thoughtful attitude toward reading and discourage the tendency to " borrow " a friend's notes.

Exercises for problem study. Neither the individual method nor the socialized recitation achieves the requisite training in facing problem situations that is so desirable for citizenship. The "open-book test" [1] is an excellent device for stimulating study and reflection, but its use is limited because of the amount of time required for directing individual study and reading the papers. Substitute exercises in the form of objective tests may be devised to induce study and thought on different problem situations. These exercises may be given to the student for study and reflection, and after their completion the results should be subjected to the test of class discussion. An exercise based on the following newspaper clipping, for instance, might be used as a preliminary to class discussion to secure familiarity with the present situation, a study of historical parallels and constitutional provisions, as well as a consideration of possible contingencies of Federal action and substitutes therefor:

The following is adapted from an article in the *San Francisco News* July 1, 1929:

> In a letter to the *News*, H. Guy, General Secretary of the Association for the Relief of Cyclone Sufferers, declared that the island of Guadeloupe wants to become a colonial possession of the United States. The status of his organization and its right to speak officially are not known.
>
> "France leaves us to our fate," said Guy, "in a formidable conflict. We say today that if France finds that Guadeloupe is a load too heavy for her, she has only to give her up to America, who will do for our country (the richest of the Antilles) the same great work of redemption she has done elsewhere in the world."
>
> "Dreadful misery" abounds throughout the island, Guy wrote. Nine months ago a cyclone killed 10,000, wounded 12,000 more, and destroyed 30,000 dwelling places.

[1] For a discussion of the open-book test see Henry Johnson's *The Teaching of History*, pages 429 ff. The Macmillan Company, New York; 1917.

To the Student:

The above comment presents a challenge to an American citizen. Below are questions indicating some of the facts you will need to know before deciding what action you think wise. In completing the following exercise, you will find different types of answers required in different statements. When blanks occur, fill in with the proper word; where more than one answer is suggested, underline the answer you think correct. Place a + sign before the number 8 if you think it is true; put a 0 if you think it false.

1. Guadeloupe is an island located between the ____ Sea and the ____ Ocean.
2. Its general direction from the United States is — (north, east, south, west, NE., NW., SE., SW.).
3. The climate of Guadeloupe is ____; its population is about ____; its principal product is ____.
4. In the history of the United States, territorial annexation due to the request of colonial inhabitants occurred in the case of — (Alaska, Hawaii, the Philippines).
5. In the case just cited, international complications did not occur because —
   the land was purchased from another nation.
   the subjects petitioned for release from the mother country.
   no other nation claimed territorial rights.
6. Such annexation of territory as was cited is legalized through Article ____, Section ____, of the Constitution.
7. This political policy is called ____.
8. Public opinion agrees as to its wisdom and expediency.
9. My opinion is that we $\frac{\text{should}}{\text{should not}}$ accede to this request.
10. My opinion is that action suggested by the letter $\frac{\text{is}}{\text{is not}}$ justified.
11. I would recommend the following procedure: _____

_____

_____

_____

RAISING THE STUDENT'S STANDARD

Through providing a goal that students can see and through furnishing a self-measuring scale that they understand, the objective test tends to raise the students' stand-

ards. Lowell considers this the most important object of the examination.

> That a standard known to the students is as important in education as in any other form of training will hardly be denied. One would not train an athlete for a mile run without having him speed over the track, giving him the time that he has made and the records for that distance. To be proficient in anything, a man must have a standard and occasionally measure himself thereby. This is, indeed, most important for the capable men. They differ more among themselves than do the average or lower tier, and hence have more need for a standard of excellence. Now in education such a standard can be set by examination, and, in fact, whether so designed or not, examinations regularly taken at the close of a period of study inevitably do set a standard for the character and extent of the work done, both for the students and for the instructors. Moreover, experience proves that college students and, no doubt, schoolboys also, will soon rise to any reasonable standard required of them.[1]

Tests sharpen a student's perception of a goal. In connection with the motivation of individual study it is significant to note the additional effort a student will put forth to reach a goal that he can see plainly. Students will work persistently at an algebra assignment because they see a certain number of problems to be completed and a definite method of procedure for the solution. The lesson is broken into steps, each complete in itself. But it is entirely too common to find students attempting to study history by indifferently reading over the pages. They find the lesson dull because they do not see events as real and they attempt to memorize words. The assignment, of course, should give life to the period studied and should emphasize a few outstanding events while subordinating details so that there will be an easily comprehended goal. If the student has in mind the ten-minute test that will come on the morrow or the review test due at the completion of the epoch, his attention

[1] *Op. cit.*, page 64.

will be directed toward the type of study that he needs. Men and women, sequence of events, causal relations, conditions of life, etc., will assume a new emphasis in his thinking. This is, to be sure, not a complete and ideal form of home study. However, it is a definite step away from the careless reading of an assignment, and the information thus acquired will furnish a basis from which the teacher may develop a consideration of problems.

Tests emphasize need for historical accuracy. It is a mistake to assume that because it is desirable to study historical situations as problems and to encourage wide reading that students should not be held responsible for any definite knowledge of the period under consideration. Hazy and undigested reading leads to fogginess of mental impressions and encourages lax study habits that are likely to result in superficial thinking in adult life. Because he knows that he may have to answer definite questions about the matter read, the student who must face an objective test attacks his study in a more alert manner. Ordinarily a student is not indifferent to a subject because it is difficult for him, but because it is dull — because it has no meaning in his consciousness. The background of events must be known, to give significance to reading or study. The learning of definite facts and the testing of this knowledge are, perhaps, small steps in the greater problem of so presenting the words of a book that they will fuse into the experience of the student. But small as this factor may seem, accuracy of impressions stands as a real step in translating words into ideas and actions. To make history seem real, men and events must leap out from the covers of a book and enter the student's experience; it requires specific knowledge about the life of other days as well as an alert imagination to give action to past events. For instance, much light is thrown on the spirit of the Revolution by the fact that the Declaration of Independence was

not issued for more than a year after the battle of Bunker Hill was fought.

Information about any historical period is significant only as it is connected with the picture of life in succeeding times. Here is a demand for the recall and reorganization of informational experience. There should be frequent summaries and reviews of facts if the passing lessons are to carry a continuous thread of interest. A test, and particularly an objective test, brings before the student the definite facts that he is expected to retain. The review with which a student prepares for a test reinforces memories and presents events in a different perspective. While the improved standard of accuracy that the student gains through preparing for the comprehensive and definite features of the new-type examination is induced by reasons different from those that influence the teacher in her concern for historical accuracy, the test is certainly justifiable as a means of obtaining drill on fundamental information.

Motivation through self-measurement. Self-measurement is an important factor in raising the standards of students. In an experiment [1] conducted in college classes for the purpose of gauging the effect of knowledge of progress on achievement, objective tests were used to measure the learning rate of different sections. The results showed that a measurement of achievement is sufficient to give those students who know their scores a distinct superiority over those students who are kept in ignorance; the degree of superiority is roughly proportional to the amount of information possessed concerning one's achievement.

A mature mind is likely to overlook the confusion of the student when he receives a poor mark on an expository paragraph that is diffuse, vague, and pointless. To the student

[1] Clay C. Ross, "An Experiment in Motivation," in *Journal of Educational Psychology*, Vol. XVIII, pages 337–346; May, 1927.

whose best effort produced the paper, the low grade means nothing in the way of correction and inspiration to try harder; he sees only that " the teacher always marks me C and she always gives James an A." Is it not true that if he had the analytical ability to see his deficiency he would not quibble so much over the low mark but would be more likely to recognize the need for greater and more effective effort? And such being the case, do teachers not owe it to the student to provide a test that he understands — a test that proves by items he may check for himself just where he ranks in learning achievement compared with his class? Students appreciate the definiteness of the new-form examination. In their own words, " When the test comes back, you know what is wrong."

True, such tests sometimes embarrass the instructor. James, of the keen mind and ready tongue, frequently trusts to his wits to carry him through without any dull adherence to sources, facts, or established authority. Again it happens that an earnest student with good power for analytical thought has difficulty in remembering details. But in the final judgment of skills is not each type of ability due its measure of appreciation for achievement? It is useless to attempt to hide from students the fact that individual differences in quality and quantity of ability exist, though of course it is most unfortunate to tell a student his estimated intelligence rating and permit him to label himself as of a definite degree of brightness or dullness. Unless teachers recognize this possible source of discouragement and show students that mere ease in learning and facility in expression do not alone determine success, many of the present generation, nurtured on tests of alertness, will develop a sense of inferiority that may inhibit effort. Character training may receive an indirect effect through showing that, though native ability is an advantage in ease of learning, other

qualities such as industry and accuracy are also influential in obtaining results.

Rewarding effort. Seeing success after effort is another reason why students like the objective examination. The very limitations of the objective test, such as its difficulty in measuring reasoning, enable it to reward industry. The student who lacks the complex skill demanded by the expository answer finds in the new-type examination a measure of his industry; while the indolent student of superior ability who cannot be roused to effort in any other way will often be shamed into exertion when he finds that he is ranking in the lower quarter or half of his class. The definiteness, the impartiality, and the range in scale are all elements bringing to the student satisfaction after effort. Interest and good-natured rivalry are often roused when students studying under different teachers compare their scores. One of the teachers who aided in the study with tests mentioned later wrote, after using Test 5, Form A (see page 157 in the Appendix), at the end of the term: "The test made a great impression on the American history classes, and their eagerness to know which class made the best score was amusing." If the same examinations, or closely alternate forms, are used a second year and the record of median and quartile points retained, students can compare their scores and those of the class with the records of the previous year.

Nothing so discourages effort on the part of the pupil as a feeling of injustice. With the essay-type of examination the teacher is often troubled by the complaint, "You marked Mary more than you did me on this question, and my answer is much longer." The attempt to point out the student's weakness requires a great expenditure of effort and time on the teacher's part, and even then she often fails to remove the feeling of injustice from the student's mind. In fact,

after reading an investigation of the reliability of subjective rating, such as that quoted in the preceding chapter, one wonders whether the complaining student is not often justified. The fairness of the new-type examination appeals to the student; he recognizes the justice that is expressed in a clearly defined goal, in impartial marking, and in success after effort. Convince a student that his mark will be fair and provide a goal that he can see and attain, and a long step is taken toward securing the effort that will raise his standard. Under such circumstances beating his own record becomes a game in which he finds real zest.

PROVIDING A MORE EFFICIENT MEASURE OF ACHIEVEMENT

Sufficient attention has already been directed to the facts that the objective test provides the instructor with a more efficient measure due to the special characteristics resulting from objectivity, and that the student appreciates the additional fairness of such scores. Other values are to be derived from the inventory test that an instructor may use at the beginning and again at the close of the term to measure the success of her teaching. The adaptability of the new-type questions to exercises that discriminate between different factors in achievement further enhances the effectiveness of this measurement. Examples of exercises suggestive for such discrimination may be found in the Working Skills Test [1] previously mentioned, and in Test 5A (on page 157 in the Appendix), where the various exercises measure knowledge of historical terms, identification of leaders, familiarity with references in collateral reading, etc. And, finally, but by no means to be ignored among the results of better measurement, is the fact that scores in the new-type tests have improved the marking system.

[1] See note on page 44.

Improving the marking system. Since there is no immediate prospect of eliminating marks, teachers should make them more meaningful, representative, and exact. The hundred or more items possible in a new-type test indicate a similar number of points on a measuring scale. However inadequate such a measure may seem in comparison with total achievement, it provides an element for the teacher's mark that is definite and impersonal and also expanded in comparison with the five-to-ten-question test of other days. Due recognition of values in the improved marking system must be accorded the rank method, which has been popularized, and perhaps made possible, through the extended scale of distribution resulting from the objective test.

The superiority of the rank method. According to the method of grading by rank, a student's mark is determined by his relative position on the scale for his class; that is, whether he is at the top, the middle, or the bottom of any scale for measuring achievement. With the light that statistical investigation has thrown upon subjective marking of essay examinations, we may smile at the naïve assumption contained in the statement that used to be printed on school report cards:

> The School Board has determined that a mark of 65 per cent or lower on the final examination [or on the teacher's record] shall be considered evidence of failure to merit promotion.

When there is no objective determination of the 100 per cent, a grade of 65 from one teacher may be equivalent to 75 from another, and so on, varying as any individual's estimate of an indefinite entity differs from other individuals' estimates. One conclusion of the Iowa study was that the marks of two teachers were not more than 20 per cent better than chance assignment.

In this investigation of examination methods in the social

studies previously mentioned,[1] the following comment is made on the superiority of the ranking method over the percentage mark:

> The differences mentioned (in marks assigned to one set of papers by two independent scorers) argue against the accuracy of the 100 per cent grading scale. Assignment of marks in terms of ranks would probably be safer, since it is easier to produce examinations from year to year which would yield approximately the same rank orders for the same pupils than it is to construct examinations which will yield approximately the same mean scores from year to year.[1]

While the rank method can be used with subjective estimates, it is better adapted to the objective scale and more equitably administered, provided, of course, that the scale is a fair one.

Percentile rank. Scores may easily be converted into percentile rank. However, it is essential that sufficient explanation be given the student to enable him to distinguish between the percentage mark to which he has probably been accustomed and the entirely different percentile rank. If the percentile-rank mark is used, the students may be given some such definite explanation as this: If your percentile score on the test is 37, that shows that your score exceeds the scores of 37 per cent of the students in your class.

The more recent books on statistics[2] furnish the formula for determining percentile rank and tables for transmuting it into a percentage score. Percentile rank is most significant when numbers in combined sections are large enough to provide expectancy for a normal distribution, and this, of course, is not ordinarily true of the small numbers in a single section or class. Graphs of achievement scores are quite an incen-

[1] G. M. Ruch and Others, *Objective Examination Methods in the Social Studies.* Scott, Foresman & Co., Chicago; 1926.

[2] Arthur S. Otis, *Statistical Method in Educational Measurement.* World Book Company, Yonkers-on-Hudson, New York; 1925.

tive for younger students, and the percentile graph [1] is one of the most satisfactory for the purpose.

Summary. Improved instruction will result from the use of new-type tests for the diagnosis of learning difficulties. Tests of that most essential skill, reading comprehension, are available with norms established to show grade reading age. Other diagnostic values will be found in tests such as the Rochester Working Skills Test, and in constructing tests with varied exercises to measure diverse types of accomplishment. Tests individualize instruction through enabling the teacher to deal with the student as an individual, to secure better study habits through the frequent use of short tests, through improved methods for checking collateral reading, and through the use of tests as study exercises that require a command of books other than the text. The new-type test raises the standard of student accomplishment by supplying a goal that is clear and definite, by emphasizing self-measurement, and by rewarding effort in study. A more effective measure of achievement is thus secured. Further results of objective testing will be found in an improved marking system based on the rank method, with scores expressed as percentile ranks if desired.

[1] Arthur S. Otis, *Universal Percentile Graph.* World Book Company, Yonkers-on-Hudson, New York.

CHAPTER FIVE

New Examination Forms Adapted to History

The technique of the new-type examination is not difficult, but it involves the observance of definite procedure. Certain types of questions have become standard; several of these are capable of much variation, and practice soon indicates which forms are best adapted to special content and teaching conditions. The leading forms are the recall, the completion, the recognition or multiple-choice, and the true-false. For those who wish to make a careful study of the different forms and the statistical method for validating tests, G. M. Ruch's *The Improvement of the Written Examination*[1] is excellent. For ordinary classroom use, however, such thorough study is not necessary, and other books that present in simple style only the more important points in test construction may be found more helpful.[2] Especially in the social studies it is important to adapt the form of question to the material used and to phrase questions to emphasize associations with other facts instead of with mere words.

Recall questions. The recall is the simplest form of asking a question and has long been in use as the " word-answer " test. This form is especially well suited to the identification of men and events, and the definition of terms. The direct recall of information is more difficult than the recognition of the correct answer when several responses are suggested, but since ease in responding to questions and in scor-

[1] Published by Scott, Foresman & Co., Chicago; 1924.

[2] J. S. Orleans and G. A. Sealy, *Objective Tests.* World Book Company, Yonkers-on-Hudson, New York; 1928.

Donald G. Paterson, *The Preparation and Use of New-Type Examinations.* World Book Company, Yonkers-on-Hudson, New York; 1925.

Charles Russell, *Classroom Tests.* Ginn & Co., Boston; 1926.

G. M. Ruch, *The Objective or New-Type Examination.* Scott, Foresman & Co., Chicago; 1929.

ing answers restricts the response to one word or phrase, the usefulness of this form is somewhat limited. In the Iowa study the recall form is credited with being the most reliable of the objective question forms and with requiring less time for correction than the other forms studied. In the new-type tests the monotonous *who*, *when*, *how*, and *where* phrases are avoided by prefacing each exercise with the necessary directions; as,

> Following are brief characterizations of certain statesmen about whom you have been studying. Write the correct name in the blank following each description:
> 1. The author of the Declaration of Independence. _____
> 2. The inventor of the telephone. _____

The completion form. The completion form is similar to the recall, and by some writers the two names are used interchangeably. A distinction will be made here by applying the name *recall* to the question that makes a unified demand for a single response, while the name *completion* will be reserved for questions asking for two or more portions or fragments of ideas within a sentence; as,

> The President's veto can be overridden by a _____ vote of _____ _____ of _____.

The recall question can be phrased so that the response blank will appear at the beginning or close of the descriptive statement, but the completion form must of necessity use blanks within the sentence. The number of words to be supplied is indicated by leaving the requisite number of blanks.

The completion exercise should be used by the social science teacher with great care and discrimination. As the response to this form depends much more on verbal memory and verbal association than do responses to any of the other forms, its use should be limited to the narrow range of content that justifies verbal memorization, as, for instance, some parts of the Constitution. Since the value of remembered

facts increases as their association with other facts becomes more complex, it is unwise to put much stress upon the recall of specific words. It is evident that the duplication of words to be memorized in a single connection makes the completion form of less value than the recall. In fact, it is well to avoid using the completion form if the question can be stated in any other way.

The multiple-response form. The simplest type of multiple response is the recognition form. In fact, this form is so easy and the chances for guessing are so high that the question is of little value unless it is used as a practice exercise or for the identification of items where memorization is not desirable. In the matching exercise the words comprising the responses to the different questions are printed in a group, and the student is directed to select from these the correct name or word for the blanks that follow the descriptive statements; as,

> Below are sentences referring to different ways of traveling. Use the words suggested at the top of the exercise to fill the proper blanks.
>
> *railroads, two-wheeled carts, horses, ships*
>
> 1. Traveling in earliest colonial times was by means of ____.
> 2. At a later period, when roads were widened, use was made of ____.
> 3. The quickest and safest way to go from Boston to Jamestown in 1650 was by ____.
> 4. Not until the early part of the nineteenth century was it possible to travel by ____.

The recognition form provides a desirable way in which to introduce students to the new-form examination; it appeals to them because of its resemblance to a puzzle that they may figure out by ingenuity. And, true enough, as soon as the better-known words are identified, it is fairly easy to guess the remaining ones. In the experiment concerned with the tests given in the Appendix, the recognition form appears in the second exercise in Test 1, Forms A and B and Test 2,

Forms A and B; but it was avoided in later tests. The tendency to guess may be decreased by including several more names in the introductory group than are needed to fill the blanks. The directions will then read:

> There are thirty names in this group, but only twenty of them are needed to fill the blanks below. On the line that follows each description write the correct name.

When all is said, however, the fact remains that the recognition form is primarily a drill exercise and of little importance for the measurement of learning.

The best-answer type of multiple response presents to the student several suggested responses, usually three to five in number, from which the best answer is to be selected by checking it or writing its number in a given space. This form has become quite popular as a result of the high rating in reliability given it by several studies based upon social science content. Multiple-choice responses have the advantage of securing objectivity without restricting answers to one word or phrase. Then, too, the element of judgment may be exercised in selecting the best answer, although the danger of personal interpretation again intrudes if responses are not confined to historical fact. Nevertheless, there is some opportunity for the introduction of questions that do not require memorized recall. The different answers permit careful shading of suggestions from the entirely false or inaccurate response through the range of plausibility to the correct answer. An example of the three-response form follows:

> After each of the following questions you will find three suggested answers. Select the one that you consider the best and write the *number* of this answer on the line following the question.
>
> Texas was annexed to the United States as a result of —
> 1. the Mexican War of 1846.
> 2. the desire for more slave territory.
> 3. the demands of American citizens who were settlers.

The difficulty in constructing alternate-response tests lies in the danger of suggesting more than one answer that is obviously true, or in suggesting alternatives whose selection depends upon opinion — an untenable solution in the social studies.

There is possible a development of this form, however, which will make for the solution of the obstacle encountered in attempting to score objectively responses to controversial issues. The device consists of stating in the suggested answers more than one argument for each of two points of view as well as one or more irrelevant items. The student may select arguments for either side of the question, but his answers must be consistent with the one point of view; as in the following:

> From the statements below select the best arguments on one side of the controversy, "The tariff versus free trade." Write the *numbers* of the arguments selected here: _____
>
> 1. A higher standard of living is made possible.
> 2. The public is ignorant of the significance of "real wages."
> 3. The policy is undoubtedly popular.
> 4. The protection of industry is essential.
> 5. Climate and natural resources should determine "home industries."

A correct answer might be either Nos. 1 and 4 or 2 and 5, but any other combination is inconsistent and should be rated zero; the inclusion of No. 3 with either of the other correct answers should be penalized by deducting the one point in error from the two points correct.

The topical-selection form. A variation of the multiple response that may be called *topical selection* seems to avoid certain difficulties in defining the best answer and to introduce a new feature. The selection question submits a group of ten to twenty statements, each of which refers to some one of three or four designated topics. To secure a keener measure of the student's power of discrimination, one or

more irrelevant statements may be included also. If this is done, the directions will read as follows:

> The following list of twenty phrases includes characterizations of each of the three topics listed as *A*, *B*, and *C*, as well as certain irrelevant statements. On the line after each topic write the *numbers* of the statements that belong to the topic. Be careful to list unrelated statements under *D* below:
>
> *A*. Hamilton's financial policy
> *B*. Gallatin's financial policy
> *C*. Jackson's financial policy
> *D*. Irrelevant
>
> *A. 1–6–7–10–11–14*
> *B. 4–9–16–19–20*
> *C. 2–3–5–8–13–17–18*
> *D. 12–15*
>
> 1. Assumption of state debts.
> 2. Creation of "pet banks."
> 3. Opposition to the national bank.
> 4. Rigid economy the keynote.
> 5. Accompanied by speculation and overinvestment.
> 6. The redemption of currency at par.
> 7. Effort concentrated on stability of national treasury.
> 8. Distribution of surplus in national treasury to states.
> 9. Formulated plans for early payment of national debt.
> 10. Tariff considered primarily for revenue.
> 11. Centralization of financial power.
> 12. Conflict with Spain over Florida.
> 13. Use of state depositories for Federal funds.
> 14. Theory of national debt as a unifying obligation.
> 15. The "Spoils System."
> 16. Special appropriations for unusual purposes becomes policy.
> 17. Tariff policy adopted for protection.
> 18. Decentralization of financial power.
> 19. Reduced military expenditures to pay national debt.
> 20. Federal policy characterized as that of thrifty individual.

On the line after *A* the pupil should write, as indicated in the form given, the numbers 1, 6, 7, 10, 11, and 14; on the line after *B* he should write the numbers 4, 9, 16, 19, 20; and so on.

The topical-selection form has the double advantage of requiring the student to know several facts about each of the

given topics and to exercise discrimination in selecting the appropriate responses for similar topics. If this form of question is new to the student, it is well to introduce it in simple form as a class exercise limited to two topics. After practice in this form, further difficulty should not be encountered by the student apart from the content of the question. For the teacher the difficulties in constructing the topical-selection form are concerned with assembling a sufficient number of descriptive statements that are at once distinctive in their application to a topic and plausible enough in reference to the group of topics to require discriminative thinking.

The true-false form. The true-false form is the best-known and most commonly used of the new examination forms. It is easy to devise and presents a needed challenge to students in training them to scrutinize plausible statements for their real import. This one purpose alone would justify the use of this form in the social sciences. Considering the frequency with which exaggerated and misleading comments on public affairs obtain currency, the student is facing a life situation when he reads a true-false statement. The true-false form has certain weaknesses, however, which must be taken into consideration if reliable results are to be obtained from the use of this form.

The element of chance is so high that no dependence should be placed upon scores unless a large sampling of questions is included; by large sampling is meant from 75 to 200 statements. In the Iowa study already mentioned, the true-false form is given the lowest reliability rating of the several forms studied.

Another weakness to which this form is especially susceptible is that of reflecting personal opinion. Particularly in the social sciences it is essential that students be encouraged to think independently. A student should never be penal-

ized for individual interpretations based on fact. Even the prefatory clause, "Check truth or falsity according to the attitude taken in this class," is unjustifiable in the social studies. So entirely is our sense of justice and truth limited by our own worlds of consciousness that a teacher is not likely to realize the force of this until she meets a true-false test in her own studying. In such a situation a teacher wrote:

> M—— gave us a true-false test. One statement was, "The amount of meaning in reading is measured by the rate of reading and the number of questions on the material read which one can answer." Our textbook states that. Yet I believe it is an untrue and vicious statement because some one phrase in what you have read may set up reverberations and echoes, may suddenly throw light upon a problem in your life, which gives a vast amount of meaning, yet you may not be able to read fast, or further, or to answer the little questions on the matter that follows the part that you have just considered so meaningful. So I had to mark the question false and take the consequences.
>
> As a result of my experience with true-false tests in which careless use of words makes different interpretations possible, so that the more you know the less are you able to mark the statement true without qualifying it, I have encouraged my students to write on the reverse side of the paper comments of interpretation. When these comments indicate a different interpretation, I accept the changed response as correct. With that safeguard to fairness, used sensibly and not in quibbling, I think the true-false becomes almost a halfway test between essay-type and new-type, teaching respect for clearness and giving practice in securing it.

Several variations of the true-false form are already in use. One of these attempts to make allowances for differences of opinion by introducing a third response symbol, such as the question mark or some similar device for indicating doubtful interpretation. Stowe has developed what he calls the A–D–U Form (*agree*, *disagree*, and *uncertain*) for use in study assignments. He says:

In A–D–U thesis-response study assignments, situations are arranged, the most important part of each of which is a thesis to which students are expected to make responses of agreement, disagreement, or uncertainty. . . .

The student's responses to all of the study theses of an assignment are made either in class or before the student begins his study of the assignment in a textbook or in collateral readings upon which the study theses are based. . . .

The student records such changes in his response as he feels ought to be made in the light of his reading and study. In the case of each response, he makes certain that he has prepared himself to present to his classmates what to him is a valid defense for his response.[1]

Another variation of the true-false form is being used by the Department of History at the University of Minnesota. After discussing certain difficulties encountered in the use of the true-false form, Krey adds:

We do, however, still use a modified form of this true-false type, one which might be described as a "double process question." The historian will recognize it as an exercise in documentary criticism. We present a series of statements either separately or in paragraph form, each statement being numbered. The students are asked to indicate whether each statement is true or false and then below, in blanks numbered to correspond to the statements above, we ask the student to indicate a proper correction. Mere chance is thereby almost completely eliminated. This type of question is not so easily marked as the more simple form, because the correction may be stated in several ways, and it, therefore, requires someone very familiar with the work of the course to grade it accurately.[2]

Difficulties in constructing true-false items. Devising true-false statements is further complicated by the condition that confronts any teacher working alone under the spur of immediately constructing one hundred or more such items that will be both intelligible and non-controversial. Ingenuity

[1] A. Monroe Stowe, *Thesis-Response Teaching in College* (Bulletin, Vol. VII, No. 4), pages 2–3. Randolph-Macon Woman's College, Lynchburg, Virginia.

[2] A. C. Krey, "What Does the New-Type Examination Measure in History?" in *Historical Outlook*, Vol. XIX, pages 159–162; April, 1928.

frequently flags, and one is tempted to fall back on circumlocution or obscurity of thought. Such procedure, of course, is not playing the game fairly. Another difficulty arises from the fact that for any true statement there are many possible untrue assertions that have not come within the observation of the high school youth. It is equally unwise to err in the other direction by presenting remarks that are obviously untrue. Ludicrous statements sometimes result from the naïve device of inserting a negative in a sentence taken from the text. For instance, the following item was included in an examination:

> Hannibal did not spend fifteen years invading Italy and hammering at the gates of Rome.

Though a negative statement is sometimes permissible, it is better to use the positive form. The double negative should never be used, as it accomplishes little more than to measure comprehension of English; as,

> The Constitutional provision that the vote may not be denied on account of race, color, etc., does not nullify state regulations refusing the vote to those who cannot read.

The type of statement that combines a well-known truth with a minor inaccuracy is confusing to the pupil, though it is occasionally valuable to emphasize that "it does not always follow," and to train students in critical thinking. Such phrasing has more value as a class exercise than as a measuring device. Qualifying words, such as *almost*, *sometimes*, *never*, etc., tend to weaken true-false statements.

It is usual to have approximately the same number of true and false statements distributed in a chance order. The easiest questions should be at the beginning of the exercise and the most difficult ones at the close.

Distinctive qualities of different question forms. In what is perhaps the most meaningful discussion of the distinctive

qualities of question forms that has appeared, Wood [1] emphasizes the fact that a change in the question form changes the significance or meaning of the question content, and that a given bit of content is not necessarily adaptable to several different question forms. In commenting on the studies by Toops [2] and Ruch [3] of the reliability of various new-type forms, Wood questions the validity of the experiment based on transforming identical content into varied question forms.

> The essential character of a question, and consequently its "statistical behavior," depends upon its content as well as upon its external form. The content of a question is often very subtle; many of the "simplest" questions are found on inspection to be compounds of many subtleties. . . . The "inner" nature of a question is more important than its "outer" form.
>
> The important thing to realize here is that objective questions arrange themselves within each external form in a hierarchy of difficulty and complexity just as the old-type tests do; and that the differences in the statistical behavior of questions in the same external form *may* be as great as, or greater than, the differences between questions in different external forms.
>
> A question involving only a simple fact of information may, with ingenuity, be put into any one of a dozen or so external forms, not counting variations in vocabulary. But with more involved and problematic kinds of questions, and even with some simple information items, the transformations can be accomplished only at the expense of varying degrees of mutilation of the essential nature of the question.[4]

To illustrate the point he is making, Wood analyzes the effect of certain transformations on the meaning of the questions involved. The first is quoted below:

[1] Ben D. Wood, "Studies of Achievement Tests," in *The Journal of Educational Psychology*, Vol. XVII, pages 1–22, 125–139, 263–269; January, February, April, 1926.

[2] Herbert A. Toops, *Trade Tests in Education* (Contributions to Education, No. 115). Teachers College, Columbia University, New York; 1921.

[3] G. M. Ruch, *Improvement of the Written Examination.* Scott, Foresman & Co., Chicago; 1924.

G. M. Ruch and Others, *Objective Examination Methods in the Social Studies.* Scott, Foresman & Co., Chicago; 1926.

[4] Ben D. Wood, *op. cit.*, pages 126 ff.

First Sample Item Used by Ruch: *Martin Luther.*

1. The reformer chiefly responsible for the reformation in Europe was ——.
2. Martin Luther was a — (missionary, pope, philosopher, soldier, reformer)
3. Martin Luther was a reformer. (*true*) (*false*)

In Nos. 1 and 2 the words *reformation*, *missionary*, and *pope* at least suggest an ecclesiastical setting for the word *reformer*. In No. 3 the student would be perfectly justified in taking the word in its ordinary meaning, and answering the question as false on the basis of Luther's intentions, his intolerance, and the dreadful immediate effects of his revolt from the Roman Church and his new political affiliations.[1]

In view of the popularity of the true-false form and the seeming ease of manipulation by the beginning test-maker whose mind is focused on the expected answer rather than on other implied subtleties, it is well to allow students to use a "modified" answer the first time a test is tried.[2]

Form for time sequence. Questions of time sequence are best phrased in the recall form, directing the student to number the events listed in chronological order or, if he prefers, to give exact dates. To avoid the injustice consequent upon scoring such a question where one error in sequence may cause the whole answer to appear wrong, it is advisable to give several groups of four or five events each. The simplest method of scoring is to allow four or five points, depending upon the number of items, for each correct group. An example of this form follows:

Indicate the chronological order of the events in each group by numbering the items 1, 2, 3, etc. Consider each group as a unit.

A. ____ The Declaration of Independence.
____ The discovery of America.
____ The Civil War.
____ The invention of the cotton gin.

[1] *Op. cit.*, page 132.

[2] For a discussion of adapting question form to different types of content see page 104.

B. ____ First steamboat on the Hudson.
____ Transcontinental railroad completed.
____ Completion of Cumberland Road.
____ Erie Canal built.

C. ____ Emancipation Proclamation.
____ First battle of Bull Run.
____ Missouri Compromise adopted.
____ Kansas-Nebraska trouble.

D. ____ Issuance of Monroe Doctrine.
____ Purchase of Louisiana.
____ California joins the Union.
____ Acquisition of Florida.

E. ____ Invention of telephone.
____ Invention of cotton gin.
____ Invention of sewing machine.
____ Invention of spinning jenny.

In marking this exercise, convenience in scoring demands that sequences correct in themselves but disarranged in serial order because of another error be disregarded. Count as correct the number of items whose serial index corresponds to the key. The correct answer for the first group above (2, 1, 4, 3) would receive 4 credits, while the partially correct answer (2, 1, 3, 4) would receive only two credits. Other ways of scoring such exercises have been devised to avoid unduly penalizing the student when a single error makes all serial numbers appear wrong. But such substitutions require a disproportionate amount of time for scoring and so complicate the work that it is difficult to turn over to a clerk. No serious injustice is done the student by giving credit only for correct answers if groups are restricted to four or five events, if the groups represent varying degrees of ease in recognition, and if the question on sequence carries proportionately small weight in the final score.

The reliability of various forms. The Iowa study of *Objective Examination Methods in the Social Studies* reports an investigation of the comparative reliability of the various

forms. In this study a sample question was phrased in each of six ways, as follows:

Recall

Eli Whitney is noted for his invention of the —.....( )

7-Response

Eli Whitney is noted for his invention of the — (1) steamboat (2) spinning jenny (3) cotton gin (4) telegraph (5) telephone (6) printing press (7) steam engine.....( )

5-Response

Eli Whitney is noted for his invention of the — (1) steamboat (2) spinning jenny (3) cotton gin (4) telegraph (5) telephone.....( )

3-Response

Eli Whitney is noted for his invention of the — (1) spinning jenny (2) cotton gin (3) telegraph.....( )

2-Response

Eli Whitney is noted for his invention of the — (1) spinning jenny (2) cotton gin.....( )

True-False

Eli Whitney is noted for his invention of the spinning jenny.[1]..( )

In this study 200 questions were cast in each of the six ways indicated above and then submitted to groups of students, 2453 in all. The investigation was well planned and carefully carried out and furnishes the best completed study with regard to the reliability and scoring of the various forms. The following conclusions are significant, but must be interpreted in connection with Wood's comment on the effect of changing the question form, and with respect to the time allowed for different forms. Time is an important factor in determining reliability, but, as the table reproduced on page 83 shows, it was not allowed in equal amount for the different forms.

[1] *Op. cit.*, page 63.

(2) Of the six types of objective examinations used in this study the recall is the most reliable. Of the recognition types, the five-response and the seven-response tests are about equal in ranking as to reliability, while the two-response and the true-false types are almost uniformly the most unreliable.

. . . . . .

(4) A comparison of mean scores arranges the six types in the following order of decreasing difficulty: recall, true-false, seven-response, three-response, five-response, two-response.

(5) The mean times required by the pupil to take the various types of examinations rank the tests in the following order as regards increasing time consumption: true-false, two-response, three-response, five-response, seven-response, and recall.

(6) For scoring, the true-false takes the greatest amount of time, the recall requires the least time, and the others are ranged in descending order as follows: seven-, five-, three-, and two-response types. This order would obtain only when the method of recording answers (by numbers in lieu of underscoring) was used.[1]

As indicated above, time is an important factor in reliability. The length of a test is the time it takes rather than the number of questions in it. If in the Iowa study the same time were allowed for all the types of questions and enough questions of each type were used so that all the time would have been spent answering test questions, many more true-false than recall questions would have been answered within the same time limit. It is probably true that in that event the true-false test would have a higher reliability — that is, when the length of the tests is the same in terms of time required rather than in terms of number of questions. The Iowa study should be interpreted with this fact in mind.

Scoring objective tests. The simplest method of scoring new-form examinations is to count the number of items that have correct responses. This is a time-saving device for the teacher; and since her tests will not be standardized, it is accurate enough for practical purposes. This method

[1] *Op. cit.*, page 88.

appeals to the sense of justice in the students and averts the antagonism that sometimes results from inflicting a penalty for the correction of statistical chance. In the Iowa investigation it was found that with " do not guess " instructions the use of the correction formula showed the true-false scores to be " badly overcorrected," [1] while the two- and three-response types were slightly overcorrected; and yet the use of the correction formula is advised by Ruch " until further study shall develop a more accurate technique." When statisticians disagree, the teacher is surely justified in using her own judgment.

The " rights minus wrongs " controversy. Since the method of scoring true-false and multiple-choice items by the " rights minus the wrongs " has had such wide acceptance, it is perhaps well to cite the reasons given by certain writers for criticizing this formula. At first the " R – W " advocates were in the majority, but later statistical studies based upon the frequency of guessing indicate that this formula inflicts a penalty that is too heavy. Chapman says that this method was the outcome of an erroneous view with regard to the operation of chance as set forth in a standard text on measurement.[2] In an investigation based upon a fifty-point test, Kohs showed that half wrong and right results occurred only once in ten chances.[3] West says:

> It appears conclusively that the "right minus wrong" method of scoring unduly penalizes subjects as a whole. This is largely due to the fact that the subjects make nearly as many errors in the items not guessed, and are given double penalty for those errors also when they should receive single penalty only.[4]

[1] G. M. Ruch and Others, *Objective Examination Methods in the Social Studies*, page 84.

[2] J. Crosby Chapman, "Individual Injustice and Guessing in True-False Examinations," in *Journal of Applied Psychology*, Vol. VI, pages 342–348; December, 1922.

[3] S. C. Kohs, "High Test Scores Attained by Sub-Average Mind," in *Psychological Bulletin*, Vol. XVII, pages 1–5; January, 1920.

[4] Paul West, "A Critical Study of the Right Minus Wrong Method," in *Journal of Educational Research*, Vol. VIII, pages 1–10; June, 1923.

The difference between the average and the individual score is pointed out by another writer:

> The effect of guessing is not eliminated in the majority of individual cases by subtracting the number of wrong answers, or a fraction thereof, from a number of right answers. Such elimination takes place only if we average the scores of a large group; but the purpose of the test is, in most cases, to find out where the individual pupil stands.[1]

The summary of findings in the Iowa study includes the following statement:

> Of the thirty comparisons of correlations of corrected and uncorrected scores, three showed the latter to be larger. All of these were under instruction not to guess. Twenty-seven cases revealed the superiority of the corrected over the uncorrected scores, and of these, fifteen were groups instructed not to guess. . . . This summary supports the conclusion that it is better to use correction for chance and instruct not to guess.[2]

. . . . . .

> The evidence collected in this study points to the superiority of the method of correcting for chance $\left(S = R - \frac{W}{n-1}\right)$ over the method of counting the score as the number of right answers. Whereas the evidence is not overwhelmingly on the side of correction, it is nevertheless sufficiently strong to warrant the continued use of such correction until further study shall develop a more accurate technique.[3]

The formula given above means that W, the number of errors, is to be divided by the number of choices minus one; that is, in the case of the true-false or two-choice question, the score is the number right minus the number wrong. In the case of the three-choice question, the score is the number right minus half the number wrong; in the case of the four-

[1] William Asker, "The Reliability of the Tests Requiring Alternative Responses," in *Journal of Educational Research*, Vol. IX, pages 234–241; March, 1924.

[2] Ruch and Others, *op. cit.*, page 74.

[3] *Ibid.*, page 88.

choice question, the score is the number right minus one-third the number wrong; and so on. In view of this difference it is not surprising to find in the usual classroom situation that the use of the formula for the true-false question may result in overcorrection for the actual amount of guessing by the individual student, although probably not for the average in a large group.

Unweighted scores are satisfactory. The practice appears occasionally of assigning different scores to different questions of a test; the more difficult the question the larger the score assigned to it. Weighting different questions according to the degree of difficulty that each presents or according to the importance of the content is a complex undertaking and one that adds little to the value of the examination. In discussing this point, Monroe says:

> Any plan of weighting adds considerably to the task of constructing the test and to the scoring of test papers. For this reason it has been proposed that the unequal difficulties of the exercises be disregarded, and that scores be calculated just as though the exercises were equal in difficulty. Although this procedure appears to be illogical, it has been found that the correlation between unweighted and weighted scores is extremely high. Charters dropped the weights from the exercises in his language and grammar tests because he felt that the differences in the scores obtained by the two methods were not sufficient to justify the added labor involved in computing the weighted scores.[1]

It is a satisfaction to the instructor with hundreds of papers to correct to know that weighting does not add to the reliability of a test and that the increased range of questioning made possible by the brief-answer test obviates the necessity for attempting to apportion different values to questions.

Deliberation and the time limit. It is doubtful whether tests in history and the other social studies should have the sharp time limitation that is customary in the earlier new-

[1] Walter S. Monroe, *An Introduction to the Theory of Educational Measurements*, page 118. Houghton Mifflin Company, Boston; 1923.

type tests. There is every reason to encourage deliberation and to discourage snap judgment in these subjects. In a time-limit test the cautious thinker who weighs the evidence on all sides is at a disadvantage when compared with a person of equal acumen who makes snap judgments. It is undoubtedly true that a conservative thinker often inhibits the immediate and ultimately correct response to a true-false statement because of a complexity of associations that do not exist for the impulsive thinker.

Sometimes slowness and depth of responses are intimately connected. Time is required in order to digest impressions and translate them into substantial ideas. "Brightness" may be but a flash in the pan. The "slow but sure" person, whether man or child, is one in whom impressions sink and accumulate, so that thinking is done at a deeper level of value with a lighter load. Many a child is rebuked for "slowness" when his forces are taking time to gather themselves together to deal effectively with the problem at hand. In such cases failure to afford time and leisure conduces to habits of speedy but snapshot and superficial judgment.[1]

The association reaction time of children is very much longer than that of adults; in the case of the latter it is from one to two seconds, whereas in the case of a child it may be four, five, or six seconds. Meumann sometimes found it as long as ten or twelve seconds. Hence the folly of requiring quick answers.[2]

Meumann found that, when children were required to give quick answers, the answer, so far as reflective thinking was concerned, was likely to be a useless answer.[3]

Another angle to this matter of limiting the time for responses is seen by Thorndike, who emphasizes the point that it is the number of right responses that is significant:

Two things are becoming clearer and surer with each year of research. The first is that there is only a doubtful value to a meas-

[1] John Dewey, *How We Think*, pages 37–38. D. C. Heath & Co., Boston; 1910.

[2] William H. Burnham, *The Normal Mind*, page 369. D. Appleton & Co., New York; 1924.

[3] *Op. cit.*, page 369. Taken from E. Meumann, *Vorlesungen* (in three volumes). Engelmann, Leipsig; 1911–1914.

urement of the general, over-all time required by a person to complete an examination in which he produces a mixture of right and wrong responses. How long it takes a person to produce fifty misspellings or mistranslations may measure something important concerning his temperament, but it does not measure his ability in spelling or French. The time required becomes significant only when the responses are substantially correct. The second is that the time required to respond at levels where the individual does have mastery is probably an even better symptom of the ability in general than we used to hope.[1]

Scaling performance. Since teachers of history have the responsibility for training students in habits that make for reflective thinking, it is advisable not to attempt to scale the performance of pupils by a time limit. The rating of different abilities that was accomplished so definitely, if not so satisfactorily, by the time limit may be secured by introducing items of greater difficulty. Such items may be details of lesser importance or associations that will be meaningful only to those students of broader reading or preparation, as, for instance, questions on our national development that require some knowledge of European history.

But although it is not advisable to depend upon the time limitation for scaling performance in social studies tests, it is quite as inadvisable to allow every student all the time he wishes. It is well to allow time for approximately one-third the class to finish all the questions and to reconsider some of them. If the teacher is trying out a new test and has not previously stated the time permitted, she may give fair warning by saying at the appropriate point, " How many can finish in five minutes? " (or " in two minutes? "). After the student has become accustomed to the new type of test and the instructor has determined a reasonable allowance, it is wise to check possible dawdling by stating a definite limit,

[1] E. L. Thorndike, "The Testing Movement in the Light of Recent Research," in *Journal of Educational Research*, Vol. XVII, pages 345–349; May, 1928.

such as ten minutes for the daily quiz and twenty minutes or longer for the unit tests.

Items possible in given time. It is difficult to estimate the number of items in the new forms to which pupils can respond in a given time, since the simplicity or difficulty of the question content affects so definitely the time required for responding. Familiarity with the various forms enables pupils to reduce the time required for a similar number of items. In the experiment with the sample tests illustrated in the Appendix, thirty minutes was allowed at first for a fifty-point test; this was soon reduced to twenty minutes. The hundred-point test was given in the usual forty-minute period, although fifty minutes would have been more satisfactory. The following time estimates are taken from Table XXXIV in the Iowa study; the questions (see samples reproduced on page 76) were used in the seventh, eighth, eleventh, and twelfth grades:

NUMBER OF ITEMS EACH TYPE PER UNIT OF TIME FOR 100 RECALL ITEMS (*Mean Times*)[1]

TYPE	UNIT OF WORKING TIME	NUMBER OF ITEMS
Recall	100 in 26 minutes	100
7-response (*g*)[1]	100 in 24.9 "	110
" (*n*)[2]	100 in 20.4 "	118
5-response (*g*)	100 in 21.1 "	120
" (*n*)	100 in 18.8 "	133
3-response (*g*)	100 in 19.1 "	132
" (*n*)	100 in 18.8 "	136
2-response (*g*)	100 in 17.3 "	141
" (*n*)	100 in 17.2 "	148
True-false (*g*)	100 in 16.5 "	151
" (*n*)	100 in 15.2 "	161

[1] (*g*) indicates the tests taken under instructions to guess.
[2] (*n*) indicates the tests taken under instructions not to guess.

[1] G. M. Ruch and Others, *Objective Examination Methods in the Social Studies*, page 84. Scott, Foresman & Co., Chicago; 1926.

Reasoning questions and the single-answer response. In discussing reasoning questions two points should be kept in mind: first, that the term *reasoning* is commonly used loosely so that the recall of complex associations is confused with reasoning; and second, that, since the essential nature of reasoning demands associations originating with the individual, the predetermined single answer of the usual objective question form demands skillful handling of the suggested responses. To illustrate the first point above may be cited the case of a writer who lists as a " reasoning test " a matching exercise in which the pupil is given such classifications as *pioneer*, *statesman*, and *inventor*, and asked to connect the correct class name with each of the proper names in a list of twenty-five persons. Even for pupils in the junior high school such activity cannot be considered as reasoning; it is not even a case of complex association. No original process is here involved; the pupil is not called upon to draw conclusions or to solve problems. The Rochester study undoubtedly gives us the best examples of thought-provoking questions so far devised for the social studies, and yet one of their teachers criticized questions such as the following as testing not ability to form a judgment but " merely the ability to recognize the right answer ": [1]

> Place a check after that phrase which you think is the wisest idea with which to complete the statement:
>
> The statement made by William Howard Taft that "the administration of criminal law in the United States is a disgrace to civilization" should be —
>
> *a.* kept out of school textbooks, because these books should not teach criticism of the government. ____
>
> *b.* made known to all, as a challenge to improve American criminal procedure. ____
>
> *c.* investigated by our officials, but other citizens need not bother about it.[2] ____

[1] *Tests in the Social Studies*, page 61. [2] *Ibid.*, page 123.

Coöperative criticism needed for tests of ideals. Historical judgments, understanding, generalizations, and abstract ideas are of prime importance in the study of history, and they are not less valuable to students because reading and class discussion have paved the way for recognition of the right answer that is not entirely original. There is need for the type of teaching that passes on the best and highest ideals of the elders, but it is essential that teachers themselves recognize the distinction involved in teaching to think in line with accepted patterns and teaching to think creatively. The process of rationalization is often so unconscious that no distinction is made between *truth* and *what one thinks is true*. This danger implies the necessity for subjecting test questions and answers, especially those other than factual, to the criticism of others. Four teachers discussed a cause-and-effect question formulated by one of them for use in a test in ancient history. The question consisted of a list of ten causes and ten effects which the students were to match. The critics insisted that no such pairing was possible; one teacher jokingly said, " I should fail on that question "; and yet when the teacher who wrote the exercise submitted it to her class, she reported the median mark to be 85 per cent correct.

Subjective judgments and the multiple-response form. The multiple-response form is admirably adapted, when sufficient space is used to develop the alternate choices, for questions demanding judgments of accepted ideals. Such a question is illustrated in brief form by the quotation relative to the "administration of criminal law" given above. But there is need to apply the criterion of truth or opinion in such questions, since the objective form may be used to cloak a response that is entirely subjective. A certain test designed for use in classes in methods of teaching the social studies presents, in the form of five multiple-choice responses, various class-

room situations from which the best and the poorest teacher-reactions are to be selected. At first thought this seems to be an objective test that induces reasoning; but since the judgments are as subjective as are the responses of different teachers to classroom situations, the reasoning does not necessarily lead to a single right response. Unless one is disseminating propaganda for a solitary " correct " method of teaching, allowance must be made for differences in individual reactions. As a matter of fact, in the use of this very interesting test with a group of college students the numerical majority agreed with the standards established by the test-makers, but it was by no means true that those students whose practice experience was most successful rated high; a nearly perfect score was made, however, by a student whose practice work and lesson planning were rated as failures by two different supervisors. Test-makers will do well to remember that when the realm of factual evidence is left, the more fully a subject is understood the more complex it appears and the more difficult of disposal by categorical means. Tests involving differences of opinion should not be used as a basis for promotion, but they are most valuable as a means of stimulating class discussion.

Examples of questions involving reasoning. In the printed tests available for the social studies it is difficult to find examples of reasoning tests that are objective. In the Columbia Research Bureau American History Test,[1] one of a group that represents the best printed tests available in the field, an approach to reasoning is indicated by a set of questions depending on similarities. For instance,

[1] Carman, Barrows, and Wood, Columbia Research Bureau American History Tests (for high schools and colleges). World Book Company, Yonkers-on-Hudson, New York; 1926.

Carman, Barrows, and Wood, Junior American History Test (for junior high schools). World Book Company, Yonkers-on-Hudson, New York; 1929.

Four of the following contributed to the foundation and development of our free public school system. Which did *not?*

There follow five phrases, not more than two of which would be recognized as part of the usual history text discussion of the topic.

The Rochester study, as was noted before, represents a definite advance in testing students' understanding of history. Even there many of the best questions would fall in the classification of "recognizing the right answer," unless it is understood that such topics have not been discussed in class. For instance, in one test, labeled "History 1–2," appears the following:

PART D. *Judgments of modern civilization based upon our study of the past.*

The Renaissance was one of the most influential factors in the development of contemporary civilization because —

____ it affected vitally practically every aspect of European life and indirectly that of the Western Hemisphere.

____ Shakespeare belonged to this period of the development of the greatest classics in the English language.

____ Raphael painted during this period the "Sistine Madonna," which became an inspiration for many later painters of many different countries.[1]

A unique exercise in this same test directs the student to write in appropriate blanks the names of the three historical characters whom he knows best, and then to select from a given list numbers indicating the various characteristics that he would ascribe to each of these characters.[2]

A wise provision for adjusting the examination to the brighter minority who are more likely to profit from thought questions is contained in the following directions, quoted again from the same test:

[1] Alice N. Gibbons, *Tests in the Social Studies*, page 136. Athens Press, Iowa City, Iowa; 1929.

[2] *Ibid.*, page 143.

The remainder of this exercise is concerned with rather difficult thought questions. You may wisely decide that you can spend your time more profitably on the other parts of the examination. You will not be marked down if you make this decision. On the other hand, if you can and do complete this exercise successfully you will receive plus credit for your work.

1. With *four* of the above periods [1] try to make an intelligent generalization concerning the nature or characteristics of the contributions of that period *as a whole*.[2]

The following examples are from unprinted tests and indicate efforts to adjust the new-type question to topics involving discrimination and summary:

The purpose of this exercise is to see how well you can interpret the connection of related events. The items listed below from 1 to 14 include happenings directly connected with either topic *A* or *B*, and one event that is distinctly unrelated to either topic. In the blank space following the statement of the topic write the *numbers* of the items related to each topic; use the space after *X* for the unrelated item. In one or two cases you may wish to assign an event to more than one topic; draw a circle around the number thus duplicated. Watch for the unrelated item. Do not guess; if you are in doubt, omit the item or items.

A. Centralization of government control. *A*._____
B. Indications of growing democracy. *B*._____
X. Unrelated. *X*._____

1. The 16th Amendment to the Constitution (income tax).
2. The 17th Amendment to the Constitution (election of senators).
3. The 18th Amendment to the Constitution (prohibition).
4. The 19th Amendment to the Constitution (suffrage).
5. Labor gains a Cabinet seat.
6. The Sherman and Clayton Acts attempt to control trusts.
7. The powers of the Interstate Commerce Commission have increased.
8. American writers gain recognition abroad.
9. Members of Congress and presidential candidates must make public their campaign expenditures.
10. Cleveland enforced the injunction against the Pullman strikers.

[1] The periods were listed earlier in the test by such suggestive phrases as: Pre-Civilized Man, Classical Greece, Europe of the Middle Ages.
[2] *Ibid.*, page 139.

11. Imperialism has become an accepted Federal policy.
12. In many states the direct primary has replaced the caucus system.
13. The number of civil service positions filled by competitive examination has increased.
14. Financial protectorates have been accepted for certain Caribbean countries.

The next exercise was designed to permit different answers to a controversial topic to be marked correct:

In the following exercise you are to select the arguments that support one side of a controversial question. Your answer may be correct, regardless of which side you support, if you select points consistently in favor of one side and *one* only. In the blank following the statement of policy that you support, put a cross; and in the blank for "arguments selected" write the numbers representing your side of the issue. There is also a blank to be used if there are arguments that you consider unsuitable or inapplicable to either side of the case, or those whose meaning you do not understand. All arguments omitted will be assumed to belong to your opponent's case.

For a policy of imperialism ____
Opposed to imperialism ____
Arguments selected _____
Arguments unsuited to either side or not understood _____

1. Constitutional justification is lacking.
2. The policy is opposed to the scheme of American nationality.
3. Such action is essential to prevent the acquisition of identical territory by another nation.
4. The right of self-determination is rendered ineffective.
5. The development of trade and commerce is favored.
6. Only sentimentality opposes the force of expediency.
7. Control of important coaling stations and naval bases is thus secured.
8. The doctrine that the "Constitution follows the flag" is inevitable but unacceptable.
9. A single government for mixed nationalities induces serious social problems.
10. American education and commercial opportunities should be extended to other peoples.
11. Commercial penetration precedes the development of an intelligent citizenry, and commerce thus obtains a sinister control over government.
12. The phrasing of the Constitution allows for this contingency.

Among unpublished tests two others are suggestive in this connection. One carries the title, "Can You Distinguish *Facts* from *Opinions?*"[1] A list of fifty statements that the student is to classify is given. Another test is under the caption, "Can You Relate Current Items in the Newspaper to an Organized Scheme of Knowledge?"[1] Directions are given for indicating the organization of fifty items that follow.

A college which is experimenting with this problem of thought questions reports:

> We have tried listing a series of events and asking the students to rearrange them in chronological, causal, or logical order of relationship. Our experience with these questions has not as yet been sufficient to justify any observations of value. Our greatest difficulties with these questions have been guarding against mere memoriter activity and working out a satisfactory scheme for grading the results. . . .
>
> Along the same line of seeking other than mere factual knowledge, we have tried hypothetical questions which call for analysis by the student or the application of generalizations and principles gleaned from similar material studied in the course. We have taken paragraphs of source material new to the students or have composed imaginary material and set a series of questions on it. This is similar to the paragraph comprehension items used in reading tests. Here, again, we are just experimenting and have as yet made no great advance.[2]

It is evident that instructors in the social studies recognize the existence of a problem in the testing of original thinking and are experimenting in a way that will ultimately bring results.

Attitude tests. The application of new test forms to the measurement of attitudes has brought little that will help

[1] The tests on Opinions and Current Items were devised by Olive Thompson Cowell of the San Francisco State Teachers College, Department of Social Sciences.

[2] A. C. Krey, "What Does the New-Type Examination Measure in History?" in *Historical Outlook*, Vol. XIX, pages 159–162; April, 1928.

2443

the classroom teacher. Nothing is gained if tests are so phrased that students give the responses that they think the instructor desires. As one critic said, "What is gained by testing platitudes?" Among the studies that have been made, Watson's,[1] in which subjects use the distinguishing words, *all*, *most*, *many*, *few*, *no*, to modify statements such as those given below, is perhaps the most significant measure of prejudice that has been worked out for the social studies:

____ Positions in the industrial world should be open to any man with the ability to fill them, regardless of racial traits.

____ Persons of pure Nordic stock are superior in intelligence to negroes.

Neumann's *Study of International Attitudes*[2] includes a good test, but it has not been published and is too long for easy duplication. Harper's investigation of the *Social Beliefs and Attitudes of American Educators*[3] is significant in its indication of the lack of unanimity in such judgments.

Attitude tests are valuable for arousing discussion and for aiding the teacher to measure the results of her instruction. There is need for a brief test that a teacher might duplicate for use at both the beginning and end of the term. To be a sincere expression of attitudes, however, such tests must be removed from any connection with term marks.

Summary. When correctly phrased, the recall type is the most valuable for informational matter since it presents a non-suggestive challenge to the memory. To insure objectivity the required response is limited to one word or phrase. The completion form should be used with caution by teachers

[1] Goodwin B. Watson, *The Measurement of Fair-Mindedness* (Contributions to Education, No. 176), page 9. Teachers College, Columbia University, New York; 1925.

[2] George B. Neumann, *A Study of International Attitudes of High School Students* (Contributions to Education, No. 239). Teachers College, Columbia University, New York; 1926.

[3] Manley H. Harper, *Social Beliefs and Attitudes of American Educators* (Contributions to Education, No. 294). Teachers College, Columbia University, New York; 1927.

of the social studies because of its dependence upon verbal associations. The multiple-response form is rated high in reliability; the following are helpful variations of this form: recognition or matching, best-answer, adaptation for controversial issues, and topical selection. The true-false form has advantages for use in the social studies, but care should be taken to avoid certain weaknesses. Question forms should be selected with due regard to the inner nature of the question. For time sequence an adaptation of the recall form is best.

The reliability of the various question forms has been rated by a recent study that used social studies content. In scoring, the "number correct" is sufficiently accurate for informal tests. Several authorities have criticized the "right minus wrong" formula for scoring true-false questions. Unweighted scores are satisfactory. The use of a sharp time limit does not seem advisable in history examinations because of the desirability for training in habits of deliberative thinking; a scaling of abilities can be arranged otherwise. The length of time required for responding to a given number of new-type questions depends to a large extent upon the difficulty of content.

The adaptation of the single-answer response to reasoning questions has developed slowly. The term *reasoning* is loosely used to mean questions of complex associations or remembered interpretations. Questions and answers, especially those of subtler meanings, should be submitted to friendly critics to avoid misleading statements. There is danger of encouraging hypocrisy if attitude tests have weight in determining the student's mark.[1]

[1] An excellent discussion of reasoning tests in the social studies, "Objective Measurement of Outcomes of the Social Studies," by Truman S. Kelley, appears in the *Historical Outlook*, Vol. XXI, of February, 1930, pages 66–72.

CHAPTER SIX

Constructing and Administering an Informal Test

Published tests in history and the social studies are concerned with the measurement of learning after the course of study has been completed.[1] What the classroom teacher needs is a series of tests that will aid instruction and serve for a summary and review of successive periods as each is completed. To adapt the new-type technique to the teaching process is at its best a problem of individual method in which experimentation in test construction will improve both teaching and testing. The steps involved in the formulation of a test will be given in some detail to aid those who are beginning such experimentation. Placing the tests in the teaching plan will be considered first; arranging the material in questions, second; and administering and scoring tests, third. The procedure described and the tests submitted should be considered as samples of what a teacher can do alone; model procedure is dependent upon the coöperation and critical judgment of an interested group or committee working together. Two worth-while reports of such coöperative experiments are available in printed form.[2]

Relating the Tests to the Teaching Plan

If testing is to be part of the instruction, a definite place should be assigned in the term plan for tests. The student should be given in somewhat permanent form, such as a mimeographed sheet, the scheme of work designed for the semester. The work plan should be divided into four or five sections, one for each month, indicating the amount of

[1] Under the classification of printed tests it has not seemed desirable to include manuals comprising questions on a single text.

[2] Alice N. Gibbons, *op. cit.* J. S. Orleans and G. A. Sealy, *op. cit.*

material assigned in topics and textbook pages for each section of work as well as for the term. A supplementary study plan giving more detail can be issued month by month. This device of a " work contract " is used with much success by the individualized study plans, for it serves as an inspiration to the earnest student as well as a warning to the careless one. Even the teacher will find that her work gains in scope from the enforced preview and plan. A place should be made for tests in the advance statement; there might well be a review test at the completion of each month's contract. While for purposes of measurement there is a decided advantage in tests that are not announced previously, the desire to use tests for instruction requires that their occurrence after regular intervals of study be known.

The mimeographed assignment should include the larger topics or issues to be studied, the pages to be covered in the text, the amount of collateral reading required, a list of suggested topics for such reading and titles of helpful books, directions for reporting on the collateral assignment, or dates for a test on such material, and the intervals or dates for the summary test. Time may be indicated in terms of semester weeks or completed topics. An excellent manual [1] based on such a plan for studying American history in high school is available, and a working guide for the year can be put in the pupil's hands at small cost. To supplement the " laboratory manual " the teacher will need to prepare only a brief sheet grouping the periods for study in units of time accomplishment and giving dates for review tests.

Emphasize study, not the test mark. A continued effort should be made to lead the students to realize that the teacher is concerned with what is learned and that testing is only one way of measuring learning. When the organization of the

[1] Howard E. Wilson, *Laboratory Manual in American History*. American Book Company, New York; 1927.

term's work is explained to the students, their attention should be called to the various factors entering into a satisfactory " mark," and emphasis should be given to the point that a single test or the series of monthly tests is only one of several elements considered in determining the final mark. Students, even in high school, can be brought to a realization that real accomplishment in history is not achieved by merely passing a test. They should be reminded, too, that the recitation is not to be considered a period devoted to drill in memorization of facts but that this hour is for discussion and study guidance. Whatever memorization is needed should be done in study time and accounted for in the monthly review test and in the brief daily quiz. An improved attitude toward tests will result if the students can be brought to recognize that their chief purpose is not to measure achievement but to aid students in more effective learning and teachers in more effective teaching.

Practice tests. Early in the term practice tests should be provided to accustom students to the new-type questions. The first may be a test dealing with ability to follow directions. It should be brief, not more than ten items, and can be corrected at once by the pupils. The second test may well be one of reading comprehension; by all means a standardized test should be used if it is possible to obtain copies. The standardized test not only saves the teacher's time in compilation but provides better scoring devices and an established measure for determining the reading age of students. Care must be used, however, in the selection of a test; not all published tests are adequately standardized, and only a few are adapted to high school use. Those listed previously will be found satisfactory, and undoubtedly there are other good tests on the market.

If necessary, a test can easily be constructed that will give a rough ranking of abilities in reading. For this purpose

select one or more paragraphs new to the students, and phrase questions as to the import, the significance of points included, and the meaning of words and phrases. By using true-false, multiple-choice, and completion forms, responses that are entirely objective can be obtained. While the value of the result is not comparable with that derived from a standard reading test, it is surprising how much difference in ability is shown by such a simple exercise. Further practice in the new test forms may be secured through brief exercises phrased on different days in each of the new question forms. Another type of test well suited to practice work is that for measuring skill in handling textbook aids and references. Then, too, the first long test in historical material may be considered in the nature of a practice test; students whose scores are low may be given an opportunity to try an alternate form after additional study, or the material of the second test may include that of the first also. While it is not advisable to continue such substitution, it is essential to avoid the fear that students so often have of tests, especially when a new factor enters the situation; and, too, the psychological effect of encouragement is worth striving for.

Samples of tests used in such an experiment as that described above are given in the Appendix. The first tests in the series were comparatively easy. The matching form, which certainly puts a very slight strain on the memory, was used in two exercises. The second test of this series was somewhat more difficult and included enough of the material of the first test so that a satisfactory mark on the second one would remove a disqualification caused by a low mark on the previous test.

Discussing test results. Corrected papers should be returned to the class for discussion and collected again at the close of the hour. Retaining the papers makes it possible to use the same test content in successive years, since students

are not likely to recall a sufficient number of items to coach classes a year later. The class discussion of questions and answers always arouses keen interest. If the class decide that a question depends on personal opinion, or if an item is considered unfair, any score for that question should be eliminated in determining the total. In addition to the emphasis on important facts that these discussions bring out, there is also a consideration of the quality of the questions: Is each a statement of fact or is there perceptible the influence of the instructor's personal bias? A continuous challenge is thus presented to the class as to the nature of evidence, and an effective motive is provided for detecting prejudice.

Simple statistical terms explained. In order that these tests may serve as a self-measuring scale of progress, simple statistical measures should be explained to the class; copy the entire distribution on the board, though of course without names. Urge each student to find his place in the upper or lower half and in the proper quarter. It is well to avoid mentioning the place or rank of any students except those few at the top of the scale. The less capable students should be shielded from undue discouragement or the ridicule of classmates, and the carelessly indifferent ones can be stirred more effectively by an individual conference. The following explanations have been found sufficient to enable an eleventh-grade class to understand a distribution:

The *distribution* shows the number of pupils that obtain each score.

The *range* indicates the difference, in number of points, between the lowest and highest scores.

The *median* is the middle score when the scores are arranged in order of size.

The *upper quartile* is the score that separates the highest quarter of scores from the next quarter below.

The *lowest quartile* is the score that separates the lowest quarter of scores from the next quarter above.

The *middle half* (or 50 per cent) of the scores is between the upper and lower quarters.

A test that is *comprehensive* provides for a wide distribution of scores, thus giving a measure for students of different degrees of achievement.

"Passing," in such a distribution as this, depends more upon the individual's place in the complete scale than upon attaining a given mark. Roughly speaking, the lower 15 or 20 per cent marks the danger line.

Small amount of time given to testing. Lest the encouragement given to students to measure their achievement by marks on objective tests seem to attach undue importance to this phase of learning, it is necessary to emphasize other types of learning in the recitation and study assignment. With the exception of the ten-minute test, the recitation hour will be used for types of learning other than memorization, such as discussions of significant events and their connection with present problems, illustrative material such as narrative sources and biography, and arguments that require the citation of evidence and serve to establish the complex association of facts with other facts and with experience. The essay test is a needed supplement to the objective form since students can be led to discuss one topic with thoroughness and with full opportunity for individual interpretation. If the recitation hour becomes a keen discussion forum and if not more than two hours out of twenty are given to examinations, there should be no danger of the students' misunderstanding the place of factual learning in the sum total of the study of history. Using the class period as a laboratory hour makes testing only one of several learning activities.

ASSEMBLING THE MATERIAL FOR A TEST

Too much emphasis cannot be put upon the fact, previously discussed, that the material used in objective tests must be important. This does not mean that all material is to be of equal importance nor that every item must be of definite significance; but it does mean that there must be sufficient material of general importance in the test to assure the average student a place in the median group. To provide for adequate measurement of the better students, items of less general significance may be included. In order to select properly the dominant events of which the test is to be a sampling, the teacher should sketch roughly all the significant happenings that the period includes. Selections for the test should be made from this material. It is not satisfactory to attempt to combine all the important events in one examination. In the first place, what seems of prime significance to one teacher will be secondary to another; again, there is danger that such a selection would result in students' learning for the recall of lists of events, men, and dates rather than of associated happenings. Furthermore, it is advisable to arrange the material so there will be two or more forms of each test for use with sections reciting at different times of the day. Then, too, in order to measure the extent of learning, there should be a range of facts from the most evident to a few of lesser importance. The greater range of difficulty gives the desired ranking of students' achievement through more complex associations instead of through time reactions.

In the first tests there is a certain advantage both to the teacher who is compiling them and to the students who are answering them in restricting the content to matters of factual evidence. This is due primarily to greater ease in constructing this type of test, but it is also true that there is a

need for factual background for the understanding of history, although this by no means indicates the necessity for memorizing lists of names, dates, battles, and so forth. It does mean, however, that the first tests should include more or less complex factual associations, the recognition of relationships, and accepted interpretations. A student-teacher prepared this question:

> New England was settled by ____.

Aside from the implied ambiguity — for the question might be answered by *Indians*, *Pilgrims*, *Puritans*, *English*, *religious dissenters*, *loyal subjects of the king*, and so on — the question is not phrased to teach desired associations. Rightly phrased questions can state and exercise desired factual relationships; as,

> The group of religious refugees who fled from England to Holland and later settled in New England was called ____.

The teacher who is beginning to experiment with test construction had better improve her skill in handling the new-type question forms in factual content and defer until a later time the effort to adjust the new forms to material that is purely thought-provoking. In writing of the beginning of the Rochester experiment, Miss Gibbon says:

> It had been decided at the first department meeting of the year to limit the practical experimentation, at first, purely to factual testing. Succeeding events proved this limitation wise. . . .
>
> By concentrating at first on the testing of the *objective of factual mastery*, the teachers learned to use the new teaching tool more quickly and were enabled more easily to see some of the values and defects of the new-type tests. There was another advantage that came out of this limitation of the experimentation to factual testing. It allowed teachers to concentrate their attention upon a better evaluation of factual details. . . .
>
> . . . These factual statements were typed and a copy given to each teacher for examination and evaluation before the group meeting occurred. At the meeting, the statements offered by any teacher had to meet the challenge, "Why do you consider that a

fundamental fact? Is that fact relevant to the attaining of any ultimate objective?" Many were discarded because they would not meet it.[1]

Apportioning material. Attention should be given to equalizing the difficulty of the alternate forms of a test and to securing the proper proportion of well-known matter and that of lesser importance. For illustration note on page 102 the way in which names of leaders were selected for Test 2A and its alternate form 2B — tests on the period of exploration and settlement. Illustrative material is taken from a series of tests prepared for an experiment in American history,[2] and samples of some of the tests are shown in the Appendix (pages 143–161).

For Tests 2A and 2B the teacher listed seventy-two names that she considered would be familiar to high school students of fairly wide reading. These names were then arbitrarily classified as the A group that every student might be expected to know without any particular study of American history; the B group, names appearing in most secondary school texts; the C group of lesser importance; and so on to the E group, with which only students of very wide knowledge of history would be familiar. It would have been far better to submit the list to a group of co-workers to obtain a composite judgment, but in the press of time the right procedure yielded to the immediate need.

It was decided to ask in each of the two alternate forms for the identification of fifteen leaders. The next step was selecting thirty names and checking them to determine what proportion of each group — A, B, C, D, and E — was represented. To equalize the two forms, the names were paired off in two groups of similar accomplishment and

[1] Alice N. Gibbons, *Tests in the Social Studies*, page 12.

[2] The tests in the Appendix and the data used in Chapter VII are taken from a doctor's thesis, *Tests in American History* (presented to the Graduate School of Education, Harvard University), by Elene Michell.

importance. A miniature scheme of distribution according to the ease of recognition was arranged, and names in the selected list were changed until the following proportion was secured:

Group	*A*	*B*	*C*	*D*	*E*	Total
Number included	1	5	5	3	1	15

The names are shown below as they were selected and paired:

Form 2A	Group Rating	Form 2B
Franklin	A	Washington
Champlain	B	Sir Walter Raleigh
James I	B	Charles I
Braddock	B	Pitt
Drake	B	Magellan
Cabral	B	Cabot
Wolfe	C	Montcalm
Gilbert	C	Stuyvesant
Andros	C	Winthrop
Anne Hutchinson	C	Hooker
Coronado	C	Cartier
Locke	D	Berkeley
Nathaniel Bacon	D	La Salle
Toscanelli	D	Henry of Portugal
Waldseemüller	E	Fox

The same procedure was followed for the question dealing with dates, but it was more difficult to find sufficient dates of equal importance for the two forms. It would have been better to ask for not more than three dates or to omit the question altogether until a later test should present a greater range. Below are the dates considered and those selected:

Dates Considered			Dates Selected		
			Form 2A		Form 2B
1492	1588	1689	1620	Settlement	1607
1496	1603	1713	1603	English history	1588
1498	1607	1733	1689	English history	1763
1500	1619	1754	1521	Exploration	1498
1521	1620	1763	1754	Attempted union	1643
1565	1643				

The total fifty points were distributed as follows:

	Form 2A	Form 2B	Revision
Geographical location	5	5	5
Identification	15	15	15
Recognition true-false	20	20	20
Time sequence	5	5	3
Cause-effect	5	0	0
Selection of topics	0	5	7

The exercise containing the cause-effect question was a disappointment; it was planned to induce reasoning, but the necessity for avoiding implications that were obscure or confusing reduced the exercise to the simple matching of consequences related in the text. In the alternate form a topical-selection exercise was substituted for cause-effect and found to be more satisfactory for inducing thought. The true-false exercise was used to include desired content that was not covered in other questions. While twenty true-false items are too few to provide a reliable measure, this difficulty was lessened since a greater number, in this case thirty items, were phrased in forms other than true-false. In formulating the true-false exercise the attempt was again made to pair each item with one for the alternate form.

Whenever possible, the test while still in preliminary form should be submitted to another teacher for criticism and for the detection of common errors referred to later as errors of poor selection, ambiguity, or faulty form. Preparing alternate forms at the same time lessens the work of construction, since items that duplicate content may be transferred from one form to the alternate. The similarity of scores made on the experimental tests by sections in different schools using both forms indicates the success of this method of pairing items. The procedure illustrated is presented not as a pattern to be copied but as an example emphasizing the importance of proportion and design in the construction of a test.

With the diversity brought to the problem by coöperative effort and the demands of varied content, other bases of approach will be developed by the teacher.

In term review tests it is essential to consider carefully the emphasis given to different epochs. For instance, a common fault in tests that have been published in United States history is overweighting with questions relating to the Civil War and events previous to it. One such test devoted 80 per cent of its questions to the earlier period; in another test designed for survey purposes in elementary schools no reference was made to any event since the Civil War except by asking for the identification of four characters. The well-proportioned test recently issued by the Columbia Research Bureau (see page 86) avoids this difficulty. The Manual of Directions carries the following comment:

> Apportionments of the questions were agreed upon in advance for the following divisions: political, social, economic, religious, and educational. Apportionments of numbers of questions were also made in advance as to the various periods of American history. Thus, 50 per cent of the questions relate to post-Civil War times. Roughly half of the questions deal primarily with what we call political material; but even in the questions which we have classed as political there is a considerable fringe of social, economic, geographical, and chronological material. About 15 per cent of the questions are primarily social, about 25 per cent are economic, and about 10 per cent are religious and educational in nature.

This is the sort of analysis that teachers have a right to expect in every published test, although the above is the only history test found containing it.

Adapting question form to content. While definite rules cannot be formulated for the use of question forms suitable to different types of historical content, much can be learned from considering the characteristics of certain forms and obvious cases of their misuse. Statistical studies, while not conclusive, place the recall form first in reliability and valid-

ity. In addition the recall is known to be an important part of the learning process, and it is safe to say this form should be used whenever possible. It is at once the simplest and most direct way to measure learning. This form has especial significance in the initial part of the learning process because it gives exercise in the repetition of truths without confusing the mind by suggesting false connections. The true-false has its own province in the realm of question forms, but its position is seriously weakened when it is used for a question that demands recall. The confusion of thinking, which the true-false is valuable for detecting, is found usually in complex ideas rather than in fundamental facts. There seems little reason for presenting in any other form simple facts that can be answered in the single phrase of the recall form. The common misuse of the true-false form for the recall ranges all the way from the absurd statements and wasted learning opportunities illustrated below to the positive danger of presenting the negative form of facts known to be commonly confused, such as " strict construction " and " loose construction." These examples of the absurd misuse of the true-false form were taken from a teacher's first test:

Samuel Adams invented the printing press. (*true*) (*false*)
William Penn was a Quaker who settled in Pennsylvania. (*true*) (*false*)

It should be remembered that a distinction is made in the use of the terms *recall* and *completion*. While both forms are limited to a single-word or phrase answer, the term *recall* is applied to the descriptive sentence that centers about a single item. Unless verbal memory is to be emphasized, as it may be in learning the Preamble to the Constitution, it is better to avoid the completion form that presents several blanks to be filled in a single definitive sentence. The ordinary identification question is admirably suited to direct recall. Time sense is best measured by the recall of sequence through

assigning serial numbers, or definite dates if preferred, to different groups of four or five events. These events may or may not be dependent upon each other according as the teacher wishes to reinforce memory or the understanding of relationship.

The true-false question. In the range of associated truths that are to be taught through the study of history there are many not adapted to questions of direct recall. There needs to be correction of many common misconceptions, some entirely false, some rather subtly twisted with partial truths. The true-false and multiple-choice forms are valuable for developing a discriminating attitude toward positive assertions. Wood refers to a question which he says "involves an absolute dichotomy, and as such is an essentially true-false, yes-no, or two-answer type of question, and not a recall or multiple-choice type."[1] As an illustration of the "inner nature" distinguishing the recall and the true-false, note the selection given previously[2] of the Iowa study's phrasing a question in seven forms. It is evident that the most satisfactory way to teach the basic fact implied as well as to measure what has been learned is to ask the direct question, "Eli Whitney invented ——," instead of needlessly confusing associations by suggesting the spinning jenny and other inventions. On the other hand consider the statement: "The Sherman Act ended the power of monopolistic trusts." Here is a definite but somewhat subtle falsity; the Sherman Act failed to accomplish that for which it was designed, but the superficial pupil who studies the event as completed needs to face the question, "Did it or did it not ——?"

The multiple-choice question. The plausible statement, especially a subtle inversion of truth, is adapted to the true-false form, while the plausible partial truths seem better if

[1] Ben D. Wood, "Studies in Achievement Tests," in *Journal of Educational Psychology*, Vol. XVII, page 137; February, 1926.

[2] See page 76.

presented in the best-answer exercise. For the large number of concepts that become attached in more or less hazy associations to other ideas, the use of the multiple-choice form offers practical training in separating popular conceptions and partial truths. For instance, if a definition of the term *free silver* is desired, no single-word or simple true-false statement fully meets the need. Here is a situation complicated by hazy thinking and partial truths. The question easily falls into the multiple-choice form:

The phrase *free silver* means—

a. the coinage without cost of all bullion brought to the mint by individuals.

b. the coinage of a large but limited amount of bullion in silver dollars.

c. the exercise by the government of certain privileges relative to coining silver in order to relieve poverty.

The discrimination required here is more difficult and more truly a teaching exercise than if one of the partial truths were phrased in true-false form.

The topical-selection question. The topical-selection question is adapted to topical summaries and to the discrimination of factors contributing to each of several situations. This form presents the advantage of attacking the problem of testing from the positive side and of demanding a discrimination in dealing with facts that goes beyond mere memorization. In the fifth exercise in Test 2B two topics are stated:

A. Economic influences that led to the exploration and colonization of America.

B. Reasons that account for the English character of our laws and institutions.

In the seven statements that follow these topics it is not difficult to select the items belonging to each topic, and yet students who are just beginning this study and have not given particular attention to these influences must exercise some independent judgment. Although the topical-selection form is capable of much variation, it seems advisable to restrict

its use to the statement of truths and to leave plausible falsities for the true-false and best-answer form. The student's power of discrimination is further challenged if he is warned that one or more irrelevant items are included in the suggested list. This form is especially valuable to determine whether students know all the facts pertinent to given topics if each of the contributing items is distinctly applicable to one situation and not to the other and if selection depends upon a recognition of relevancy to the particular topic. Material dealing with topics that have received special attention, such as the weaknesses of the Articles of Confederation and the compromise issues in the adoption of the Constitution, may justifiably be intermingled and given to students for organization.

What is a good question? The amateur test-maker is confronted with a situation that is at once an obstacle and an advantage — the fact that the only way to find out which questions are really good ones is to try them with actual students. The time spent in revising tests is counterbalanced by the advantages gained in having an improved test available for repeated use. For the purposes of the informal classroom test the statistical measurement of reliability and validity is not essential. A fair evaluation of the comparative difficulty of items may be obtained by tabulating the number of correct responses, errors, and omissions for each item. Although questions are selected to cover individual differences in learning, it is nevertheless advisable for the teacher to scrutinize items of importance upon which as many as 50 per cent of the class have failed. The implication may be that the question is a poor one or that additional teaching is needed. Another method of evaluating tests is to compare success for a particular exercise or item with the total score for an individual or with the teacher's subjective estimate of the student. Wood calls attention to the " in-

verted question," upon which the poor student makes a better mark than the good one; subtle suggestions that pass over the head of the dull or careless student sometimes trip up the one who is more alert. For instance, in the following question taken from a test the student whose knowledge is limited to the English colonies has an advantage over the one who is also familiar with the adventures of the French convict sailors:

The men who came to live in the colony were —
a. rich members of the company.
b. poor people who wanted to make a better living.
c. sailors from the ships.

Unless care is exercised to eliminate the inverted questions, they may defeat the whole purpose of examination. The teacher who checks examination scores to detect weaknesses in various items will be rewarded not only with better tests but with a sharper recognition of the fundamentals in learning history.

Difficulties to avoid. Common errors in test construction fall in three groups: errors resulting from (*a*) poor selection, (*b*) ambiguity, (*c*) faulty form.

(*a*) *Poor selection.* In the selection of material no error is more common and no danger more pernicious than the failure to make the content of a test representative of material that has significance for the understanding of history. The inexperienced teacher who constructs a test with the textbook open is particularly liable to such stupidities as these:

The Aztecs were a French tribe. (*true*) (*false*)
The people traveled to America in _____ in 1492.
The compass helped the people — (*select one answer*)
on water in walking
on horseback on a train
The third President to be assassinated in office was _____.

Repetition of content is common; one test excellent in other regards has six different items hinging upon the lack of religious toleration in the Puritan colony. Such repetition not only defeats the comprehensiveness that is one of the chief assets of the new-type test, but it often results in giving away the answer to a second question; as,

> The period from 1830 to 1850 accomplished little in the way of social reform. (*true*) (*false*)
> A social reform movement that had many adherents in the days preceding the Civil War and that led more or less directly to that conflict was concerned with ____.

The matching form is especially susceptible of giving away answers by means that are not intended. Certain grammatical constructions will designate to the alert student which word in a given list is needed, as when a plural verb or the article *an* precedes a blank, when an adjective is called for and only one occurs in the list of suggested words, or when an antithesis is presented and only one word in the group could complete it.

A fault connected with the selection of responses is the suggestion of alternatives that intelligence alone, quite aside from historical judgment, may recognize as poor, mediocre, and good. For instance, a test asks the student to determine which of three summaries best fits an accompanying paragraph. The reader can see without the context, and so could the bright high school student, which of the following is best:

> *A*. The Crusaders learned many things from their enemies, including new foodstuffs, new ways of dressing, and manners.
> *B*. The Crusaders acknowledged that their heathenish foes were more highly civilized, finer men than themselves
> *C*. Upon returning home, the Crusaders introduced into Europe a few important customs which they copied from the Turks.

(*b*) *Ambiguity*. Teachers who have themselves taken examinations and tried to guess which one of several inter-

pretations the instructor was thinking about will sympathize with the child who answered the query, " Columbus sailed for ____," with *India*, only to find his answer marked wrong because the teacher expected *Spain*. The demand for a single word as the answer is justifiable only when the sentence is so clearly phrased that no other interpretation can be given it. The following exercises might have had several correct answers supplied for each blank:

Daniel Boone was a ____.
France found no ____ but furs in America.
Sir Walter Raleigh tried to build ____ in America.

An ambiguous question is as inexcusable as a stupid one, and probably more common. Mention has been made of the confusion resulting from the use of double negatives. Statements that are merely misleading should be avoided, although the need to train for the detection of the *non sequitur* conclusion justifies the purposeful use of such clauses. The following statement, designed to find out whether students recognized the existence in the early colonies of racial elements other than English, is probably too misleading to be justified:

Before the Revolutionary War the settlers in the thirteen colonies were English or the descendants of English. (*true*) (*false*)

The " tricky " question is often a poor one that attempts through vagueness to test discrimination. Care should be used with the qualifying words, *any*, *almost*, *usually*, *some*, *never*, and *none*, which make a falsity too evident or entirely destroy the point of an otherwise true statement. The following is an amusing example of a teacher's attempt to change a sentence in the text by inserting *not* and *any:*

The invention of the compass was not of any help to the people of Europe in learning more of other lands. (*true*) (*false*)

(*c*) *Errors in form.* Among common errors in form are indefinite or confusing directions, the failure to include a sufficient number of items, using question forms with a "hit-

or-miss " relation to content, and placing the more difficult items at the beginning of an exercise or a test. Directions should precede each new exercise; they should follow an accepted mode and be so clear that no confusion will result and that responses will be written in the designated place only. Since the reliability of the new-type question is in large degree determined by the length of the test, it is important that a large number of items should be included. This is especially important to remember when constructing true-false tests or other forms where suggested responses enable chance to become a factor. The best authorities agree that a true-false exercise should have not less than 75 items, and preferably from 100 to 200; it has been suggested that final examinations for use in colleges should have upward of 500 items. The need for such long tests is modified by the classroom situation when tests are for teaching purposes, by the occurrence of frequent tests as opposed to a single final examination, and by the use of other forms that lessen the chance element so prominent in the true-false type of question.

ADMINISTERING AND SCORING TESTS

Directions should be provided for teachers and for students who may use the test. The directions to teachers should include some such statement as the following:

DIRECTIONS FOR ADMINISTERING THE TEST. This test is one of several that will be given at intervals of —— [or at the completion of each one of the various units decided upon]. Since several classes may use the same test, it is essential that conditions be as nearly the same as possible. You are asked to observe the following conditions:

Seat the pupils alternately. (If that is not possible, see that each student is seated according to a definite seating chart that you may use later if it seems advisable to check certain cases of identical errors.) Distribute the tests face down. Because we may wish to use this form again, be careful to collect extra copies or sheets

that may have been distributed, or spoiled sheets when new ones are asked for. If the students are not placed in alternate seats, see that each desk is provided with a sheet of blank paper. Ask the students to use this for covering the page upon which they are working, explaining that this is not only a protection to them but will save annoyance to neighbors whose glance inadvertently might turn their way.

When you are ready for writing to begin, announce that there is to be no talking and that no questions are to be asked (unless in mimeographing a paper is blurred), make note of the time (in writing), and direct students to turn their papers over as you read to them the directions at the top of the test [or the directions given you for students]. You are asked to give your whole attention to conditions in the room, ventilation, etc., and while it is advisable to remain seated at your desk — students often complain of distraction when a teacher walks about — do not busy yourself with clerical or other work that will absorb your attention and do not leave the room while the examination is in progress. Unless the honor system is established, the responsibility is yours to see that there is no opportunity for dishonesty. Keep the attitude of the room serious but avoid any sharpness or tension that will suggest strain or anxiety to the students.

See that no papers are retained longer than —— minutes from the time of starting. [*Or:* Collect all papers when all but two or three students have finished.] Collect immediately the papers of any students who finish earlier. Allow no talking until every paper has been collected.

After you have corrected the papers, send to Room —— a copy of the scores listed in tabular form (names are not essential) in order from the highest score at the top of the page to the lowest at the bottom. While median and quartile points will be calculated for all sections together, it should be remembered that a normal distribution of scores cannot be expected in the comparatively small group represented by a single class.

Directions to the student might be as follows:

The test is one of several you will have this term to aid you and your teacher to see what progress you are making in history. On your desk you should have two sharp pencils and a sheet of blank paper which you are asked to use as a cover for your work. You

will be allowed —— minutes for the test and will find the time sufficient if you work briskly. [*Or:* More questions are included than you are expected to find time to answer. *Or:* You will be allowed ample time, but do not waste it, since the papers will be collected when all but a few have finished.] It will be wise to go through the test rather rapidly at first, answering the questions you find easiest. If a question seems hard, do not guess; skip that question for the time being, and when you have finished the other items, use the remaining time to think about the questions you skipped. Do not talk, and do not ask any questions unless there is an error in typing that makes the paper difficult to read.

While it may seem unnecessary to use such definite directions for classroom tests, the business of taking examinations is simplified by the establishment of a habit-pattern that prevails whenever a test is announced. And it may serve to break down some of the nervous tension so unfortunately associated with tests to have students become accustomed in the everyday classroom quiz to the same formal procedure. Once the procedure is established, it will be unnecessary to repeat the detailed instructions.

Completed form. New-type examinations may be written on the board under a protective roller map, or the questions may be dictated. But either practice is less satisfactory than the use of mimeographed or printed sheets. In a large school the mimeographing should be taken care of by the commercial department, or the printing department if there is one. Adequate precautions should be taken to guard the copy from inquisitive students; those who are studying stenography should have a heightened sense of the confidential responsibility of the secretary. Copy intended for student typing must be letter-perfect with exact spacing, margins, blanks, etc. In the smaller school the teacher who types can make her own stencils, and if school paper is supplied, copies will be run off by a commercial establishment at small cost. Stencils may be preserved and used repeatedly. The

sample tests in the experiment described later were mimeographed, and while the scope of the questions was limited by the plan of covering fifty items in two pages of typing, the resulting simplicity and brevity of set-up were appreciated by the students.

The appearance of test material on the typed or printed page influences the student's attitude of zest for an interesting problem or aversion toward a dull exercise. While the primary reason for advocating the use of different forms in a single test is to adapt the form of question to the nature of the content, there is also the opportunity to appeal to individual differences in students and to provide variety. The use of several exercises, each employing different question forms and variously spaced blanks for answers, tends to relieve the monotony resulting from a large number of similar questions and aids in keeping the student's interest fresh until the end of the test. The crossword puzzle would not have attained such popularity had it not presented large, blocked spaces in a diagram with pleasing variation of line. The same material set up in fine type with run-on lines would have been endlessly dull. The features of pleasing make-up may be achieved to some extent even in the mimeographed copy prepared by a classroom teacher. Legibility of print, clearness and simplicity in directions as well as in questions, diversity of exercises, and the consequent variation in length of lines, frequent double spacing, and an occasional bit of facetiousness slipped in as an absurd response — all such devices tend to relieve the strain of examinations and decrease the student's fatigue.

Margins and spacing. Ease and rapidity in scoring are increased by provision for marginal responses. The spacing thus required avoids crowding more than one response item on a line. There has been a tendency to provide blanks in the left margin for true-false responses and in the right for

recall and other forms. Ease in scoring would place all blanks in one margin. The arguments for the left margin are the convenience of placing the answers immediately next the number for the item, and the assurance of a regular margin line. The right margin is in line with the natural movement of the eye in reading, but it requires for the typist a preliminary lining off of answer space before typing is begun and a repetition of numerals adjacent to the space to guide the eye to the proper blank.[1] Care should be taken to see that certain exercises, such as the topical-selection, where the latter portion is dependent upon reference to the earlier part of the exercise, are typed on a single page and not separated with part at the bottom of one page and continued on the next. Using a double space after every fifth true-false item aids the eye in following the score key, and also makes the exercise easier for the student to read.

It is important both for the reliability of the new-type test and for convenience in scoring to construct all questions so that answers are entirely objective. However, the teacher who prepares her own tests may provide such variation from purely objective responses as desired, provided she is consistent. If questions requiring answers in sentences are included, it is better to place these questions on the last page of the test. This permits the student to use additional space for his answer without disturbing the alignment of the scoring key, and also makes it possible for the instructor to divide each paper, giving the objective portion to a clerk for scoring.

Safeguarding the honest student. If the suggestions given as to margins and spacing are followed, the completed test will be in a form that is comparatively easy for a student to see on a neighbor's desk. Not only are answers easier to copy from an objective test, but because scores are so definite,

[1] In the case of the published standardized test it is preferable to use the right margin for the answers in order to take advantage of the easier scoring.

students lose confidence in the test results if there is opportunity for dishonesty. To avoid annoying the honest student and tempting the dishonest one, care should be taken to guard against copying answers. Whenever possible, alternate seating should be arranged; such a plan is essential for an examination of any importance. If the classroom is not large enough for this, a cover sheet of blank paper should be provided for each student to use in covering his work. The teacher will need to be alert, for in a test that rates students through comparative achievement, it is doubly important to guard the honest students from unfair competition.

Scoring papers. The score key should be spaced so that the responses on it will be aligned with the student's answers when the key is placed beside the examination paper. The directions for each exercise should call the student's attention to the line where the answer should be written. The use of plus and zero signs gives greater speed in correcting the true-false question than does the use of letters. The minus sign is unsatisfactory because of its insignificant size and the ease with which it can be changed to a plus sign after the corrected papers are returned to students. The multiple-response form may be scored more quickly if the alternate responses are lettered *a*, *b*, *c*, etc., or numbered 1, 2, 3, etc., and a line or pair of parentheses is provided on or in which the student is directed to write the letter or number designating the response he selects. Directions for the various question forms may be seen in the sample tests in the Appendix.

In scoring tests, use a colored pencil and mark each item in the margin exactly opposite the response but not penciling through it. For convenience in re-checking, an accepted scheme of marking should be used; a satisfactory plan is to use a straight line for a correct response, a cross for an error, and a zero for an omission. Adding the score for these three items and comparing with the total for each exercise is one

check on careless errors in scoring. If the test is scored according to the number correct, it is not so essential to distinguish between errors and omissions. The occurrence of clerical errors is so common that there should always be provision for re-checking scores. In cases other than the term-end test this checking will be taken care of if tests are returned to students for the class discussion that is so valuable a part of examinations. At this time students should be warned not to make a mark of any sort on their papers, and to have no pencils or pens on their desks; they will readily respond to the teacher's desire to be fair, and during the discussion errors in scoring can be brought to the teacher's attention and corrected.

In the ten-minute quiz, papers can be exchanged and corrected by the class. Difficulties are prevented in this case by having answers copied upon a duplicate sheet of paper which is given to the teacher before answers are read and marked by the class. Correcting papers by exchanging is not satisfactory for a test of any length, although a few selected students can be of help in scoring the usual classroom test. The most satisfactory way to use their assistance is to have answers read aloud while two or three students each correct a paper, scoring all tests on page 1 until the set is completed, then all tests on page 2 until the set is finished, and so on. Unless tests are returned to the writers, as in the period of class discussion, it is necessary to have a second person re-check each paper. The number of clerical errors that slip into scoring, even by responsible adults, is amazing.

Those who have participated in the testing movement in the past few years will find these directions unnecessary, but the beginner in teaching or testing will find it advantageous to establish certain methodical reactions in regard to scoring. Papers may then be filed with the assurance that the record is a clear one if for any occasion a re-checking is called for.

Form subordinate to content. Perfection of form in a new-type examination will not cover poor selection of material. The construction of an objective test demands much more than the adaptation of content to certain question forms. Because of the larger sampling of content made possible, a greater responsibility is put upon the teacher for the elimination of unimportant details and the presentation of a significant sampling that bases achievement on the recall of meaningful associations. The new-form tests are susceptible to grave abuse in that they may become fairly exact instruments for measuring verbal memory. Teachers need constantly to remind themselves of how much they have forgotten of the subject matter learned in youth; even more, they need to realize how much they can afford to forget and still pass as educated. The memory should hold the significant facts that give the common element so necessary for the transfer of knowledge about government into adult intelligence in regard to political affairs. But such facts should not be learned as isolated items. They should rather be learned as part of a connected whole related to daily experience. An educator who is not directly concerned in teaching the social studies says:

> The selection of material is the first task toward making history educational. Instead of covering ground, we must have an intelligent goal and drive toward it. There is only one key to selection, and that is the present and, so far as human wisdom may judge it, the future which is flowing out of the present. Those actions in history which are still living and which are thrusting their forces into the future are the substance of the future.[1]

Summary. (1) *Relating tests to the teaching plan.* Through the use of a work sheet containing plans for the term's study, a definite place should be made for tests. The fact should be emphasized to the student that it is learning and not

[1] Edward C. Sisson, *Educating for Freedom*, page 168. The Macmillan Company, New York; 1925.

measuring with which the teacher is concerned. Practice tests should be used to familiarize students with new question forms, and the results of tests should be discussed with the class. A simple explanation of statistical terms will enable each to find his place on the scale of distribution in order to evaluate his work in comparison with that of the class.

(2) *Assembling material for a test.* For the beginner in test construction it is advantageous to restrict early tests to factual material. The problem of apportioning subject matter in the test is of outstanding importance. The test must carry an abundance of significant material, and the various epochs in time development as well as the different types of social science content must receive due attention. Question forms must be adapted to the nature of the questions. Several criteria may be used to determine what is a good question, and whenever possible the test should be submitted to someone else for criticism. Certain outstanding difficulties that recur in test construction should be guarded against.

(3) *Administering and scoring tests.* Directions for administering tests should be so clear and definite that coöperating teachers will meet no difficulties. In mimeographed papers the appearance of the completed test deserves attention in order to secure a page that is clear, easy to read, and satisfactory to the eye. Alternate seating of students should be arranged in order to protect the honest student. In scoring papers there is need for great care, and re-checking should be provided for. Certain formal devices bring speed and accuracy to the task.

CHAPTER SEVEN

An Experiment with Uniform Tests in Four High Schools

Certain questions that will occur to those considering experimentation in tests will be answered as far as is possible from the data secured in a single study. This, of course, is not conclusive data, but it presents material for thought. The study [1] was based on the use of a series of tests in four high schools, two in an Eastern state and two in the West. In the Western schools, referred to as W and X, the tests were a definite part of the teaching plan; in the two Eastern schools, Y and Z, the tests bore no relation to the teaching scheme except that whenever possible they were administered at the close of appropriate units of study.

The tests in the series were given at intervals of about four weeks during the fall term and at approximately the same time in the several schools. The students in the Western schools were in "required" classes of United States history in the eleventh grade, while in the Eastern schools similar classes were in the twelfth grade. There was kindly coöperation in administering the tests, but the teachers in School Z were very skeptical as to whether tests of information could measure anything worth teaching to high school seniors. The situation there was also complicated by the fact that only the college-preparatory classes followed the chronological order of study upon which the tests were organized. These facts should be kept in mind in interpreting the data presented later in the chapter.

The questions to be considered are: (1) What advantages are there in using uniform tests in sections studying the same period of history? (2) To what extent should the score on an objective test influence the teacher's mark? (3) Can the

[1] Elene Michell, *Tests in American History*. (See footnote on page 101.)

teacher be judged from the class median? (4) What achievement in history can be expected from the student of normal intelligence? (5) Is there a common basis for measurement in history?

OF WHAT VALUE ARE UNIFORM TESTS TO A DEPARTMENT?

If tests are used as part of the instruction for different sections studying the same term of history, values will be apparent additional to those for which tests are commonly used in schools. The coöperation among teachers in planning to cover requirements previous to the test, the discussion of aims and bases for selecting material, the acceptance of a distinctive core of minimum essentials for each term of study, and equalizing the standards of different teachers are among the definite gains. But in discussing the advantages of uniform tests there should be no quibbling over the fact that their use can have a detrimental effect on the teaching. However, the abuse of an opportunity is not a good argument against its value. If uniform tests are to be used effectively, there must be a realization both by the supervisor and the teachers that such tests can and should measure only the minimum essentials for a term of study, and that a teacher's most valuable service to the department lies in the individual contribution that each teacher gives to her classes in addition to covering the content agreed upon. Neither the supervisor nor the teachers can afford to lose sight of the fact that uniformity holds no value so great as the individual instruction of a capable teacher.

Minimum requirements equalized. However, after individuality has received due recognition, there is left a significant advantage in uniform tests. Since they insure the consideration of minimum essentials, they tend to adjust demands made upon pupils in similar courses by the different teachers. Inequality of this sort is common in large schools

where the demand of one teacher is quite rigorous and of another entirely too lax. Coöperation in test construction necessitates the discussion of standards, with a resulting value far greater than the equalization of requirements. In many cases where a supervisor would hesitate to criticize or dictate standards, the conferences attendant upon preparation for an approaching test will solve the difficulty. This measure is especially helpful in organizing a new department or in introducing new teachers to an established standard.

Varying standards within a department. The teacher who requires the highest type of study is frequently the one who marks most rigidly. In one of the schools where this experiment was in progress, a teacher said to the investigator, "Is there anything in your study that will show that the other teachers have a lower standard of marking than I have? Pupils often ask to be transferred from my classes in order to get higher marks, and the principal considers me too severe." This instructor, who will be called "A," and her colleagues, "B" and "C," were each teaching sections in United States history. Tests had been given at approximate intervals of a month and at the close of the term; scores [1] were averaged for test forms 2A, 2B, 3A, 3B, 4A, 4B, and 5A, with the following result:

TABLE 1

SCORES MADE ON TESTS BY PUPILS OF THREE DIFFERENT TEACHERS

	TEACHER A	TEACHER B	TEACHER C
Median	60	50	42
$Q_1 - Q_3$	52–70	42–62	26–56
Range	40–78	27–75	11–71
Number of students	28	103	122

[1] Tests 1 to 4 inclusive were composed of 50 items; they were scored on a basis of number correct multiplied by 2 to give a total score based on a possible 100. Form 5A included 100 items and was scored on the basis of number correct.

It was evident that Teacher A was maintaining a higher standard of achievement than the other instructors; in fact, her students were the only ones in this school whose scores were comparable with the records of three other schools. But a study of term marks given by the three teachers showed that A's students were receiving a smaller proportion of high marks than the students of the other two instructors. Table 2 shows the percentage of marks that each teacher allotted to the various letter grades.

TABLE 2

PERCENTAGE OF LETTER GRADES GIVEN BY THREE TEACHERS

TERM MARK	*A*	*B*	*C*	*D*	*E*	NUMBER OF STUDENTS
Teacher A	15	37	48	0	0	26
Teacher B	31	35	32	2	0	103
Teacher C	4	59	32	4	0	122

While the numbers in Table 2 are not large enough to justify the assumption that the marks will approach the normal distribution, it is plain that Teachers B and C are giving more marks of *A* and *B* than would be earned in the ordinary class. In consequence the students of Teacher A are penalized, as far as marks are concerned, for being in her class, though from the standpoint of good teaching, the investigator would have placed Teacher A's class first. Had there been any disposition to study the uniform test scores, the variance of standards might have been adjusted. But the attitude of these teachers was that tests of information were no adequate measure of their instruction. Of course, tests imposed from without are never so acceptable to a department as tests that the teachers have constructed themselves.

WHAT INFLUENCE SHOULD THE RATING ON AN OBJECTIVE TEST HAVE ON THE TEACHER'S MARK?

While the examination score should never be allowed to supersede the teacher's estimate of a pupil's achievement, it is nevertheless desirable to offset the subjective mark with an objective rating. This is a protection to the teacher as well as to the pupil. But since the ranking of students secured by the new-type tests is such a definite and convenient rating, there is danger of putting too much reliance on the result of a single test; if objective scores are to weigh in determining the final mark, a series of such tests should be arranged.

The evidence that the present study has on the relation of test scores to the term mark is indirect. It is significant particularly in that it shows inconsistencies when no objective element was considered. The teachers' marks, available only for the two Massachusetts schools, were quite independent of the test scores. Of the four teachers whose marks are compared with their students' averages for the test series, only two indicated any desire to see the objective scores, and even for these two the scores for the term review test were not ready before the teachers' marks were given out. Tables 3 and 4 show respectively for Schools Z and Y the teacher's term mark compared with the average test score for each student. In School Z, Forms 2A, 2B, 3A, 3B, 4A, 4B, and 5A were used at intervals of about a month each; in School Y, all but the first form were used. All tests were given in the fall term and as near the same time as was consistent with having a "period" test follow the unit of study on which it was based. While many of the marks are not consistent with the test scores, and perhaps should not be, there is such glaring inequality in the cases of Teachers B and C as to seem significant, especially when considered in

connection with the class medians as shown in later tables. Note for instance the fact that these two teachers gave a mark of *B* to students whose scores ranged from as low as 25 to 28 per cent on the series to as high as 64 to 71 per cent. Even though the marks were independent of test scores and the teachers had no voice in determining questions, it is to be expected that a mark of *B* given to a high school senior in United States history would discriminate among students whose knowledge of historical fact is as low or as satisfactory as this range suggests. Furthermore, if the position taken here is sustained — that effective learning and reasoning in history are dependent upon knowledge about historical events, and that historical perspective is dependent upon specific information — then high marks were given to students ignorant of essential learning.

A mechanical marking of students according to objective scores is not to be desired; it would ignore too many important factors. It can be seen from Table 4, which gives the record of School Y, that in spite of inconsistencies there is probably a reasonable approximation between the objective score and the teacher's estimate of achievement. No student averaging less than 61 made a positive *B*. The student scoring 83 evidently deserved no more than the *B* given him, since his IQ was one of the highest and he could have achieved this score with little effort. The test scores of the *A* students ranging from 66–80 indicate that the teacher's mark included other types of pupil performance besides an information test.

From the data presented here it would seem that within certain broad margins high marks should not be given to students who are noticeably low in objective test scores and that a minimum amount of historical information should be required of high school students who make a passing mark in history.

TABLE 3

Scores on History Tests B, C, D, and E Averaged and Compared with Teachers' Term Marks. School Z

	Teacher A	Teacher B	Teacher C
Teacher's Mark[1]	A (15%)[3]	A (31%)	A (4%)
Test Scores[2]	60 68 74 76	35 48 59 64 44 50 59 67 45 54 59 69 39 56 59 71 47 53 60 71 48 58 61 74 48 58 61 75 43 55 47	45 49 54 62 68
Teacher's Mark	B (37%)	B (35%)	B (59%)
Test Scores	49 59 63 64 64 65 67 68 72 78	28 42 51 64 29 43 51 31 43 54 34 45 54 35 45 56 35 47 57 39 47 57 40 48 57 40 48 58 41 48 59 41 49 60 41 50 61	25 29 33 40 49 57 25 29 33 43 49 57 25 29 33 43 49 58 26 30 33 43 49 59 26 30 34 44 50 61 28 30 35 44 50 62 28 30 32 47 50 62 29 31 32 48 51 64 22 35 41 48 51 64 38 39 41 48 51 66 39 41 41 48 53 71 39 54 54 53 53 71
Teacher's Mark	C (48%)	C (32%)	C (32%)
Test Scores	40 59 47 60 48 60 50 64 54 66 57 67 59	27 38 43 49 57 29 38 43 54 59 28 39 43 55 59 31 39 44 55 59 32 40 44 56 61 33 40 49 54 62 34 42 49 37	21 31 38 44 46 55 25 31 39 43 45 57 26 32 40 41 50 58 28 32 41 44 51 60 28 33 41 48 53 28 34 43 48 54 31 37 43 42 54
Teacher's Mark	D (0%)	D (2%)	D (4%)
Test Scores	None	43 47	11 20 32 38 38
Teacher's Mark	E (0%)	E (0%)	E (1%)
Test Score	None	None	42
Total Number of Marks	28	103	122

[1] Letter grades are the teacher's term marks.
[2] History test scores are given in numerical rating.
[3] The percentage in parentheses indicates percentage of students assigned the letter grade by each teacher.

TABLE 4

SCORES ON HISTORY TESTS C, D, AND E AVERAGED AND COMPARED WITH TEACHER'S TERM MARKS. SCHOOL Y

	Teacher F
TEACHER'S MARK	A (10%)
TEST SCORES	66 74 76 77 79 80 73 75 77 77 80
TEACHER'S MARK	B (26%)
TEST SCORES	61 64 68 70 73 77 61 64 57 70 74 79 62 64 69 70 76 83 62 64 69 72 77 64 66 70 73 77
TEACHER'S MARK	B – (11%)
TEST SCORES	41 47 55 59 62 64 45 53 55 60 63 57
TEACHER'S MARK	C (30%)
TEST SCORES	35 47 53 51 63 70 37 48 54 55 63 64 43 48 54 56 66 42 49 55 59 66 45 52 55 60 68 46 52 50 61 58
TEACHER'S MARK	C – (13%)
TEST SCORES	32 37 45 47 50 55 62 33 40 47 47 52 61 67
TEACHER'S MARK	D (9%)
TEST SCORES	43 50 54 55 59 46 53 55 56 67
TEACHER'S MARK	E (0%)
TEST SCORES	None
TOTAL NUMBER OF MARKS	107

TABLE 5

MEDIAN SCORES OF DIFFERENT SECTIONS ON THE SAME SERIES OF TESTS

SCHOOL	TEACHER	TEST 1				TEST 2				TEST 3				TEST 4				TEST 5			
		MED.[1]	Q_1[2]	Q_3[3]	NO.[4]	MED.	Q_1	Q_3	NO.	MED.	Q_1	Q_3	NO.	MED.	Q_1	Q_3	NO.	MED.	Q_1	Q_3	NO.
X	P	64	56	74	61	66	54	76	58	64	52	74	60	64	56	74	62	56	58	76	59
	Q	74	62	78	78	68	58	76	69	58	50	68	75	63	50	70	72	52	40	62	73
	R	70	62	80	54	74	68	78	56	68	60	74	57	70	62	78	46	66	62	74	44
Y	F									62	55	73	99	67	54	76	99	52	44	64	101
Z	A					62	54	70	32	64	56	72	31	58	56	64	29	50	44	62	30
	B					58	44	66	78	56	52	70	87	48	38	60	86	36	29	43	143
	C					58	48	68	87	44	30	52	90	52	40	60	94	36	28	44	150

[1] Median score.
[2] Lower quartile.
[3] Upper quartile.
[4] Number of students.

CAN THE TEACHER BE JUDGED FROM THE CLASS MEDIAN?

The scores made by the students of a given teacher signify little about the quality of instruction if only one section or one objective examination is considered. Not only is the matter of individual differences operative, but every teacher knows that two sections taught concurrently may react in totally different ways. However, the scores for a series of tests or for different sections usually show a characteristic trend which becomes significant if the same characteristics of distribution reappear in different sections. Table 5 gives the record of median scores for the same tests for sections taught by seven teachers in three schools. The intelligence rating, summarized in Table 6, showed some little difference in the abilities of students in the different schools; but the school with the highest IQ's made the lowest test scores. Such a comparison of median scores as is given in Table 5 should not be considered significant of the quality of teaching unless it is reinforced by less mechanical testimony. However, it is worthy of notice that Teachers B and C, whose marks as studied previously seemed poorly distributed, have sections rating much lower than other teachers.

TABLE 6

IQ's in Four Schools, Based on Group Intelligence Tests

	School			
	Z	Y	X	W
Median IQ	122	101 [1]	115	117
$Q_1 - Q_3$	114–127	94–108	106–120	110–124
Range	83–135	78–129	78–134	86–135
Number of Students .	98	97	227	117

[1] The record for School Y is based upon scores derived from the Terman Group Test A, while the records for the other schools are derived from the Dearborn Group Test C. This difference may account to some extent for the seemingly low rating of School Y, since a corresponding difference was noticed when students in School X were rated with both tests.

Varying exercises to measure achievement. Another type of class measurement is made possible by designing different exercises to measure various skills or kinds of knowledge. For instance, in the term review test used in this study, the first exercise dealing with the meaning of historical terms had in every class correct responses to more than half the items. No other exercise ranked as well, although the recognition of true-false statements, when scored number correct, ranked second. A record of the percentage of correct responses in each exercise is given in Table 7.

TABLE 7

PERCENTAGE OF CORRECT RESPONSES IN EACH EXERCISE IN TEST 5A ARRANGED ACCORDING TO CLASSES OF VARIOUS TEACHERS

EXERCISES	TEACHER						
	A	B	C	D	E[1]	F	G[1]
I. Historical terms	.73	.60	.61	.64	.53	.73	.64
II. Time sense	.33	.24	.26	.31	.28	.46	.36
III. Recognizing true-false	.65	.53	.55	.60	.51	.68	.70
IV. Identification	.64	.24	.28	.33	.34	.58	.48
V. Constitution	.52	.34	.34	.55	.31	.41	.48
VI. Selection of topics	.53	.34	.27	.36	.26	.52	.35
VII. Collateral reading	.12	.03	.01	.07	.04	.06	.17
VIII. Analogies	.37	.26	.30	.30	.25	.40	.24
Number of students	31	110	113	117	60	103	58

[1] The students of Teachers E and G had not taken any other test in the series.

This table reads as follows: The students of Teacher A answered correctly 73 per cent of the items in Exercise I relating to historical terms, 33 per cent of the items in Exercise II relating to time sense, etc.; there were 31 students who took the test in classes taught by Teacher A.

By far the most difficult query, judged on the basis of the smallest number of correct responses, was that which asked for the titles and authors of five collateral references. The fact that the range of correct items is only from 1 to 17 per cent in this question is a significant comment on the scattered quality of the collateral reading required by different

teachers. The low percentage of correct responses to VI, the topical-selection question, indicates either a lack of careful and organized thinking with reference to these topics or ignorance of them. While this form of question may be difficult at first, for most of these students it was the third test in which they had used this form. The percentages of the classes of Teachers E and F fall no lower than the classes of Teachers B, C, and D although these students had no previous practice with this form of question, while the usual superiority of the classes of Teachers A and F is noticeable in this question.

Consideration of a series of tests or records made by several sections seems to indicate a certain level of achievement maintained by the different teachers or by their sections. If this characteristic level persists over a length of time, it may be taken as an indication of the teaching level of the instructor. When certain exercises devoted exclusively to one type of work — as, for instance, the use of reference books — show a very low percentage of correct responses, there is an indication, other things being equal, that this type of study is being neglected.

WHAT ACHIEVEMENT CAN BE EXPECTED IN HISTORY FROM THE STUDENT OF NORMAL ABILITY?

One of the puzzling problems that every teacher faces is the extent to which a student's achievement is influenced by his ability in relation to effort. School W [1] was selected for study in connection with this problem because its record on the tests was highest and hence indicated a more effective expenditure of effort. As shown in Table 6, School Z had the highest ability rating, although the seemingly low intelli-

[1] "W" is the Union High School located at Redondo Beach, California, the only high school in a city of about 15,000. Miss Imogene Spaulding is head of the Social Science Department, and it was under her direction that W's record was made.

gence level of School Y is due in some degree to a different standard of measurement — the Terman test always rating lower than the Dearborn.

TABLE 8

AVERAGE SCORES MADE IN SCHOOL W IN SERIES OF HISTORY TESTS BY STUDENTS OF SIMILAR INTELLIGENCE LEVELS [1]

IQ	AVERAGE SCORES ON HISTORY TESTS	NUMBER OF STUDENTS
126–135	66 69 74 75 77 80 84 86 89 67 72 74 76 79 82 84 88 92 68 72 75 76 80 82 84 89 94	27
116–125	44 58 63 67 71 72 73 74 76 77 79 83 85 88 54 59 66 69 71 72 73 75 76 77 82 83 86 55 62 66 69 71 72 73 76 76 78 82 85 87	40
106–115	50 52 59 61 64 66 67 69 73 75 77 77 82 51 53 60 62 64 66 68 71 74 75 77 78 93 51 59 60 62 64 67 69 72 74 76 77 78	38
96–105	60 62 64 67 68 70 71 76	8
86–95	42 43 44 57 59 61 66 75	8
Total		121

[1] Students are shown in five groups, determined by an 10-point interval on a scale ranging from an IQ of 86 to 135. Scores derived from Dearborn Group Intelligence Test C.

The records for School W have been arranged in two tabulations to show in the first, Table 8, the different scores made by students of similar IQ, and in the second, Table 9, the different IQ's of students making similar test scores. Table 8 is based on an IQ range of 85 to 135 points, divided into intervals of ten points each; thus in the lowest group ranging from 85 to 95 are found test scores of 42, 43, 44, 57, etc. The same data appear in Table 9 but from a different point of view; the basic scale here is the range of scores on the history tests, averaged for each individual, and extending from a score of 40 to a possible 99. Within each of the five intervals of 12 points each are shown the IQ's of the students

making such scores. For instance, the lowest scores, from 40 to 51, were made by students whose IQ's were 86, 91, 92, 111, 114, and 115.

TABLE 9

IQ's of Students in School W Making Similar Average Scores in Series of History Tests [1]

Average Scores	IQ's	Number of Students
88–94	113 126 127 128 128 133 135	7
76–87	97 111 116 118 120 122 126 127 129 106 111 117 119 120 124 126 128 131 107 114 118 119 121 124 126 128 132 107 115 118 120 121 125 127 129 134	36
64–75	87 103 108 110 113 115 118 119 122 126 128 95 105 108 110 114 115 118 120 122 126 128 99 106 108 111 114 116 119 120 124 127 131 100 107 108 111 114 117 119 122 126 127 103 107 109 113 115 117 119 122 126 127	53
52–63	90 94 105 108 108 109 117 119 120 92 100 107 108 109 110 118 119 124	18
40–51	86 91 92 111 114 115 121	7
Total		121

[1] Scores for each student were averaged for Tests 1, 2, 3, 4, 5, and ranged in five groups on the basis of 12-point intervals of score from 40 to a possible 99 (94 was the highest average obtained).

Certain implications in this record are of interest. Each interval in the scale (based on IQ's) has test scores above and below the median, 73, although the two lowest groups show few scores above the median. Only one student with an IQ above 106 has a score lower than 50, while the highest ability group has no score below 66. The two lowest groups have no score above 76.

Although this study of one school offers insufficient data for proof, yet the following assumptions seem justified for the test forms 2A, 2B, 3A, 3B, 4A, 4B, and 5A:

(1) No student with the IQ of a normal individual should average lower than 50.
(2) No student with a high IQ should average lower than 65.
(3) A student in the lowest fifth of the scale of IQ's will not average above 75, even with much effort.
(4) A student with normal intelligence may with sufficient interest and industry make a high mark.

Considered in connection with these assumptions, the following tabulation of medians from different schools is significant.

TABLE 10

MEDIAN SCORE AVERAGED FOR SERIES OF FOUR TESTS IN FOUR HIGH SCHOOLS

SCHOOL	Z	Y	X	W
Median	44	64	64	73
$Q_1 - Q_3$	36–59	54–68	54–70	66–77
Number of students . .	231	107	214	113

The intelligence quotients cited in this study were obtained from group intelligence tests, and hence should be considered as merely comparative estimates. Since the median of intelligence estimated for high school students is from ten to twenty points higher than the median estimated for the voting population, teachers of the social studies can find some encouragement from the study, which shows that even the student of mediocre ability can be expected to gain a foundation of historical facts upon which to build his political reactions.

TABLE 11

MEDIAN SCORES FOR CERTAIN OF THE TESTS

TEST FORM	MED.[1]	Q_1–Q_3	Q[2]	RANGE	NUMBER OF STUDENTS	SCHOOLS INCLUDED
1A	69	58–80	11	32–96	244	W, X
1B	68	58–78	10	30–90	124	W, X
2A	68	52–76	13	10–96	278	W, X, Z
2B	62	58–70	6	2–96	441	W, X, Z
3A	62	52–72	10	6–99	466	W, X, Y, Z
3B	60	52–72	10	6–92	239	X, Y, Z
4A	56	44–66	10	12–88	128	X, Y, Z
4B	64	54–72	9	2–98	186	X, Y, Z
5A	46	34–60	13	4–96	976	W, X, Y, Z
5A[3]	58	46–68	11	24–96	493	W, X, Y

[1] Tests 1 to 4 inclusive were composed of 50 items; they were scored on a basis of number correct multiplied by 2 to give a total score based on a possible 100. Form 5A included 100 items.

[2] Q is the semi-interquartile range and is half the difference between the upper and lower quartiles. It is a commonly used measure of variability.

[3] The second tabulation of scores on test 5A was made to show the difference when the records for School Z were eliminated.

IS THERE A COMMON BASIS FOR MEASUREMENT IN HISTORY?

In the preceding chapters certain measurements for skills were suggested and note was taken of several studies relating to attitudes. In this study of uniform tests in four high schools no evidence was collected as to the measurement of either of these factors. But as to the measurement of that part of knowledge comprising historical fact, there is abundant evidence. This evidence tends to show that there is a common body of knowledge covered in the usual high school test in history that can be measured by uniform tests without injustice to students of different schools or localities. To substantiate the conclusions to be drawn, note the following:

1. Tables 5, 10, and 11 show that a median score below 50 is not common.
2. Table 5 shows that in schools X and Y the greatest range in medians occurred in the term review test 5A; it was 52–66. The median range in the third school, Z, was 36–50.
3. Comparing the combined median for the schools with the medians of the various sections for each of the different tests, in only two cases (Teachers B and C) did the individual median fall more than 4 points below the combined median.
4. To substantiate the point that sections taught by these two teachers, B and C, are below average, note that their records as shown in Table 5 are consistently lower than those of any other teacher; and in Table 12 that

TABLE 12

SCORES MADE BY CLASSES OF DIFFERENT TEACHERS IN TEST 5A

SCHOOL	TEACHER	MED.	Q_1–Q_3	RANGE	NUMBER OF STUDENTS
X	P	56	48–66	30–78	59
X	Q	52	38–64	30–82	73
X	R	68	62–76	40–92	44
X	S	54	42–68	26–82	30
Y	F	52	44–64	14–88	101
Y	G[1]	44	36–54	24–70	57
Z	A	50	44–62	14–74	32
Z	B	36	29–43	14–68	143
Z	C	36	28–44	12–74	150
Z	D[1]	44	32–56	4–76	105
Z	E[1]	32	24–44	12–64	55
W[2]		67	56–73	38–96	131

[1] The sections listed for Teachers G, D, and E did not take any other test in the series.
[2] School W tabulated all records for three teachers as one.

these two sections are among the three lowest in the group of fourteen teachers.

Two conclusions seem justified: (1) The median score in these tests should be above 50; (2) the record of School Z is very different from that of the other three schools.

The original premise was that instruction in the history of the United States requires a more or less vicarious participation in problem situations of the past as they are related to present conditions; that for effective citizenship this learning cannot be separated from information about the development of political ideals and the fundamental structure of government; and that since learning in history implies the ability to reason about social and political problems, such reasoning is seriously limited if there is ignorance of the development and organization of the state. This study has shown that in three out of four schools, classes in the history of the United States knew 50 per cent or more of the facts asked for in the series of tests used. Therefore it seems reasonable to assert that, other things being equal, the students in these three schools are better equipped to lead in civic life and to reason from a basis of fact about civic issues than the students of the fourth school, whose test record was much lower.

It is not assumed that ability to pass an objective information test is evidence of having received good instruction or of possessing civic ideals; unquestionably a knowledge of fact may exist without stimulating effective behavior. But it can be asserted that, given two teachers who are equally strong in their desire to inculcate right attitudes and ideals, that one is the better teacher who can combine such an aim with instruction as to the nature and development of government. It is not without significance in this connection that, of the classes of the fourteen teachers listed in the tabulation of the review test, the median marks of nine of these classes

showed from 50 to 64 items correct out of a possible 100; and that four of the five teachers whose classes fell below this mark (rating 32, 36, 36, and 44 items correct out of a possible 100) were in a school that does not consider a knowledge of facts about government essential for citizenship. It cannot be denied that teachers differ in ability and effectiveness of instruction; but from the records in the case there is no implication that the nine teachers whose class scores were highest were ignoring ideals and attitudes. The contrary is probably true, especially when it is considered that the classes of the four heads of departments come within this group.

This discussion has been confined to the values and uses of objective tests in the teaching of history. Important as testing is, however, it is conceded without hesitation that other points of method hold broader significance. While such matters of method are not within the province of this discussion, it is fitting to express appreciation of the significant movements developing in connection with several experiments in progressive teaching of the social studies. As progressive methods shift emphasis upon the type of historical material studied, leadership in the determination of test content should be taken by the teachers themselves.

The most valuable purpose of tests is for the improvement of instruction, and the most satisfactory tests are those developed by teachers of the subjects concerned, teachers who have some knowledge of the new examination technique. To emphasize the importance of such experimentation, no better evidence can be offered than the following excerpts from the 1927 report of the Committee on History and the Other Social Studies submitted to the American Historical Association with plans for a five-year survey: [1]

[1] A. C. Krey (Chairman of Committee), "History and Other Social Studies in the Schools," in *Historical Outlook*, Vol. XVIII, pages 110–124; March, 1927.

Improvement of Tests. The old-fashioned examination has been under fire for some time both as a device of instruction and as an accurate measure of achievement. More important is the fact that psychologists have devised methods of testing and measuring results of instruction which promise to prove useful in the teaching of the social studies. The need for the most accurate testing devices possible arises in connection with nearly every division of this study. While relying still upon sensitive and well-trained teachers to furnish the most important gauge of results, it is desirable also to enlist whatever aid objective tests may offer. There is need of such help in the determination of objectives, organization of content and grade-placement, as well as in the formulation of examinations. For this reason it seems desirable to make the improvement of tests in the social studies a major division of the proposed study. Some of the aspects of testing which should receive attention are:

1. A study of older methods of test and measurement. . . .
2. The present use of the new-type tests in the social studies. . . .
3. The formulation of tests to measure more subtle values. A rather casual survey of tests thus far developed indicates that the new-type examinations have not reached out far beyond the measurement of specific information. The efforts to test "moral judgment," "thought," "time-sense," and "place sense" are, as their titles indicate, pioneer efforts which must be greatly refined before they can be of practical use. Very little has, as yet, been done to measure such values as the acquisition of important social points of view, attitudes, and ideals; skill in the handling of material; discrimination between fact and opinion; recognition of authoritative work; and the awakening of new or deeper interests, vocational or recreational.
4. Testing the validity and the reliability of new tests. . . .
5. A study of the function of tests. . . .
6. Recommendations on the use of testing methods both new and old. . . .

Finally it must always be remembered that the whole problem is primarily a responsibility of the teachers of the social studies and that the psychologists can give only advice and counsel. Even after several years of the best effort, it is not to be expected that all the values of the social studies will lend themselves to precise mathematical measurement.

Summary. Questions that might occur to those considering experimentation with tests are answered as far as is possible with the data secured from a study of uniform tests administered in four high schools.

(1) *What advantages are there in using uniform tests in sections studying the same period of history?* Coöperation among teachers planning to cover requirements previous to the test, the discussion of aims and the bases for selecting material, the acceptance of a distinctive core of minimum essentials for each term of study, and the equalization of standards among different teachers, these are among the definite gains.

(2) *What influence should the scores on objective tests have on the teacher's marks?* A mechanical marking of students according to objective rating is not desirable, but within a fairly broad margin high marks should not be given to students who are weak in historical information, nor a passing mark to those who fail to give evidence of knowing a minimum amount of historical information.

(3) *Can the teacher be judged by the median score of her classes?* Considered for a series of tests or for several sections, there seems to be a level maintained by different teachers or sections. If the aims of the sections and the abilities of students are the same, this teaching level may be considered indicative of the instructor.

(4) *What achievement can be expected in history from the student of normal intelligence?* In the series of tests used, no student of normal IQ should score less than 50, while, with sufficient effort, the average student can attain a high score. Those in the lowest fifth of the scale of ability will not score above 76, even with much effort.

(5) *Is there a common basis for measurement in history?* The evidence of this study relates only to the measurement of that portion of historical knowledge that is established

fact. The median scores in the series of tests as well as the medians in the review test show a similar achievement for three of the four high schools, the fourth school being one that did not accept factual learning as an objective. While similarity of scores in an objective test does not indicate similarity in ideals and attitudes, it must be granted that, given two teachers who are equally strong in their desire to inculcate right attitudes, the one who can combine such an aim with instruction in the nature and development of our government is achieving better teaching, and her students will be better able to reason from a basis of historical fact about civic issues.

The program of the American Historical Association stresses the importance of experimentation for the purpose of studying the function, nature, and value of tests in the social studies.

APPENDIX

Illustrations of Tests Used in an Experiment in Four High Schools

The following explanation was sent the teachers who participated in the experiment:

The purpose of these tests is not to measure the comprehensive work of history teaching, such as acquaintance with literature, formation of individual judgments, or the establishment of right social attitudes. These tests emphasize the basic foundation of facts essential to impartial thinking and avoid matters of individual opinion and controversial questions.

Such objective tests need to be supplemented occasionally by an hour in which the student is required to write fully on one comprehensive topic, such, for instance, as "The Articles of Confederation: the events occurring while they were in force, the form of government they permitted — including powers and weaknesses — and the political and economic situation leading to the adoption of the Constitution." Students need to be encouraged both in study and in examinations to strive for a broad "overview" of important topics. While the ordinary high school student finds it difficult to select the salient facts so that he can present a subject adequately in forty minutes, yet he needs occasionally to grapple with the task of covering the broad sweep of a single topic and to recognize that mastery of the subject does not consist in "stating four weaknesses of the Articles." Study based upon the use of a topical outline where the main points are suggested by the teacher will prove a valuable training in seeing a subject comprehensively and developing it succinctly. A test on a topic such as the above where the answer is required in outline form with subtopics in some detail is an excellent supplement to an objective test and also a check on the verbose student who pads his examination paper with unsupported opinions.

Test 1, Form A is designed partially as a practice test, and for that reason some of the items are repeated in Test 2, Form A; if you prefer, Test 1, Form A may be omitted altogether. There will be four tests in this series to be given at intervals of about four weeks. With the exception of the term review test each period

test has an alternate form for use with sections reciting at different hours of the day.

The fact that a class may not have studied certain of the items occurring in the tests should not disturb the teacher. In order to provide a scale for rating the various degrees of attainment, the tests have been planned to include a minimum number of minor items with a maximum amount of important facts. A test in which five members of your class make 100 per cent gives you no comparative rating for these five students; hence, each test should contain a few questions that only the unusual student can answer. Furthermore, there is no predetermined passing percentage for these tests; this should be ascertained by each teacher for her own class by determining above what mark the highest quarter of the class stands, below what mark the lowest quarter is found, and so on.

It is not desired to have a sharp time limit for these tests. It is suggested that the instructor use twenty to thirty minutes, as she sees fit. The term review test should have additional time — at least fifty minutes — and if possible should be given to all sections at the same time.

NOTE. In order to show the conditions of the experiment described in Chapter VII, these tests are presented as originally used with only minor changes. The scores given for Tests 1, 2, 3, and 4, with their alternates, were based on *twice* the number correct, thus changing the score on the fifty-point test to a percentage of 100.

TEST 2, FORM A

PERIOD OF EXPLORATION AND COLONIZATION

1492–1763

Name.....*Grade*.....*Date*.....

I. Give the geographical location by continent for each of the following places:

1. Canary Islands ____ 2. Acadia ____ 3. San Salvador ____
4. Cape Horn ____ 5. Fort Necessity ____

II. The following phrases characterize certain leaders in history; write the correct name on the blank preceding each description. There are more names than you will need to use.

Ptolemy; Montcalm; James I; Queen Anne; Benjamin Franklin; Cabral; John Locke; Nathaniel Bacon; James Wolfe; Toscanelli; Anne Hutchinson; Coronado; Champlain; Sir Francis Drake; Waldseemüller; Braddock; Sir Humphrey Gilbert; Sir Edmund Andros; Thomas Paine; Magellan.

____ 1. The writer who first used the name "America."
____ 2. The discoverer of Brazil.
____ 3. An Italian scholar whose map Columbus used.
____ 4. The first governor of Canada.
____ 5. The English advocate of the "divine right of kings" theory.
____ 6. Leader of the English forces at Quebec.
____ 7. Founder of an unsuccessful colony, who was himself lost at sea.
____ 8. An explorer in the Southwest and in the Colorado Valley.
____ 9. The first English commander to circumnavigate the globe.
____ 10. Leader of an uprising against an unjust colonial governor.
____ 11. A religious leader driven from Massachusetts Bay Colony because of a difference in religious opinions.
____ 12. An English philosopher who wrote the feudal constitution for the colony of Carolina.
____ 13. The British general under whom Washington served in the Seven Years' War.
____ 14. The royal governor of Massachusetts when James II fled from England.
____ 15. A colonial leader who was the author of the Albany Plan of Union.

III. Write the correct date before each of the events listed below or number the events in the order of their occurrence. Use either type of answer, but do not combine the two forms.

____ Plymouth was settled.
____ Queen Elizabeth's reign ended.
____ William and Mary accepted the "Bill of Rights."
____ The first circumnavigation of the globe was completed.
____ Colonial delegates at Albany discussed the Plan of Union.

IV. Some of the following statements are true and some are false. Put a + sign before those that are true and a 0 before those that are false. Do not guess; if you do not know the answer, leave the space blank.

____ 1. The English flag was the first to circumnavigate the globe.
____ 2. The New England colonies were founded on the principle of religious toleration.
____ 3. The earliest explorations in the interior of North America were made by the Spanish.
____ 4. The wealth of silver taken from South America built up local industries in Spain.
____ 5. During Elizabeth's reign Protestants from the continent found a refuge in England.
____ 6. Canada was settled by French Huguenots.
____ 7. The physical character of the Atlantic coastal plain determined the location and extent of the early colonies.
____ 8. Independence in local government resulted from charters granted to trading companies.
____ 9. The purpose of the New England Confederation was to arrange trading privileges.
____ 10. Geographical isolation removed the colonists from conflicts due to the mother country's participation in European wars.
____ 11. Massachusetts developed independence in government during the Commonwealth in England.
____ 12. The revolution of 1688 and its increase in parliamentary power resulted in a decrease in restrictive navigation laws in the colonies.
____ 13. The Dominion of New England was a voluntary association for protection against the Indians.
____ 14. Undemocratic class distinctions existed in colonial society.
____ 15. Before the end of the first century of colonization, shipping and shipbuilding assumed economic importance in the colonies.
____ 16. All laws passed by colonial legislatures were subject to the approval of England.

____ 17. All residents of the colonies might vote for members of the colonial assemblies.
____ 18. As a result of the Seven Years' War, France regained the colonies she had lost to England.
____ 19. The presence of racial stocks other than English was noticeable in the thirteen colonies.
____ 20. By 1763 good roads and vehicles were common, and transportation between the colonies was popular and cheap.

V. Below are two groups of statements. The first states causes, and the second results or related conditions.[1] Each one of the results is to be paired with its related cause by writing the *number* of the result in the blank space preceding each of the numbers 1 to 6. For instance, before statement 1 write the number of the related result 9.

Causes

____ 1. The printing of books began about the middle of the fifteenth century.
____ 2. James I believed in the "divine right of kings."
____ 3. The early explorers sought gold and the "Northwest Passage."
____ 4. England finally defeated her old sea rival, Holland.
____ 5. The first English colonists were unwilling to plant crops.
____ 6. In the century that followed the end of the civil war in 1485, England was laying the foundation of future greatness.

Results

7. The New World was known for a century before permanent colonization north of Mexico was effected.
8. The defeat of the Spanish Armada marked the rise of a new world power.
9. Columbus was influenced by an account of Marco Polo's travels.
10. The name of New Amsterdam was changed to New York.
11. Under the first Stuart, England persecuted Dissenters.
12. Terrific suffering and loss of life occurred during the Virginia "starving time."

[1] The form here used is that of the original experiment. An improvement would be effected by including more than the required number of results so that the process of pairing would become one of discriminative selecting rather than of matching similarities.

TEST 2, FORM B

PERIOD OF EXPLORATION AND COLONIZATION

1492–1763

Name.....*Grade*.....*Date*.....

I. Give the geographical location by continent for each of the following places:

1. Madeira Islands ____ 2. Louisburg ____ 3. Strait of Magellan ____
4. Cape of Good Hope ____ 5. Fort Duquesne ____

II. The following phrases characterize certain leaders in history; write the correct name on the blank preceding each description. There are more names than you will need to use.

Drake; Vasco da Gama; Sir William Berkeley; Thomas Hooker; George Washington; John Winthrop; Peter Stuyvesant; Henry of Portugal; Magellan; George Fox; La Salle; Jacques Cartier; Montcalm; William Pitt; Sir Walter Raleigh; William Penn; Charles I; James II; John Cabot; Cabral.

____ 1. An English adventurer who sent the first expeditions to Virginia.
____ 2. The English king whose arbitrary rule caused the Puritan emigration.
____ 3. The first governor of the Massachusetts Bay Colony.
____ 4. The discoverer of the continent of North America.
____ 5. A famous European who spent his life studying navigation.
____ 6. The commander of the first fleet to circumnavigate the globe.
____ 7. A Frenchman who sailed up the St. Lawrence seeking India.
____ 8. Discoverer of the Ohio River, explorer of that vicinity; sailed down the Mississippi to its mouth.
____ 9. A minister who moved his congregation from Boston to found Hartford.
____ 10. A colonial governor who hated free schools and feared education.
____ 11. The young colonial surveyor who mapped the route followed by Braddock.
____ 12. A governor of New Amsterdam who had trouble with settlers in Connecticut.
____ 13. The leader of the French forces at Quebec.
____ 14. An English statesman whose policy enabled England to win the Seven Years' War.
____ 15. The founder of the Society of Friends.

III. Write the correct date before each of the events listed below or number the events in the order of their occurrence. Use either type of answer, but do not combine the two forms.

____ The first permanent English settlement in America was made.
____ France ceded Canada to England.
____ The Spanish Armada was defeated.
____ Vasco da Gama sailed around Africa to India.
____ The Virginia House of Burgesses was organized.

IV. Some of the following statements are true and some are false. Put a + sign before those that are true and a 0 before those that are false. Do not guess; if you do not know the answer, leave the space blank.

____ 1. The "English sea-rovers" was the name given to officers in the royal navy.
____ 2. The Puritans were intolerant of different religious beliefs.
____ 3. The Spanish made permanent settlements in North America.
____ 4. Spain made most of her New World conquests before the defeat of the Spanish Armada.
____ 5. The French colonists in Quebec were religious refugees.
____ 6. The colonial policy of France was more liberal than that of England.
____ 7. In extent North America is about twice the size of Europe.
____ 8. The purpose of the Mayflower Compact was to guarantee equality of opportunity.
____ 9. The New England Confederation was an alliance to stop smuggling.
____ 10. Many white laborers in the colonies were kidnapped from the streets of London.
____ 11. All royal governors were appointed by the king.
____ 12. The colonists found that geographic isolation was effective in protecting them from conflicts due to European Wars.
____ 13. The religious leaders of the Puritan colony were opposed to democracy in government.
____ 14. Life in colonial times was much as it is today.
____ 15. Agriculture was the only industry possible in the colonies.
____ 16. The governor of a proprietary colony could veto acts of the colonial assembly.
____ 17. Only a small percentage of the colonists could vote.
____ 18. The Seven Years' War brought an indecisive victory to England.
____ 19. The inhabitants of the Thirteen Colonies contained a noticeable representation from nations other than England.
____ 20. Travel between Jamestown and Boston was cheap and popular when these settlements were young.

V. Read carefully Topics *A* and *B* that follow this paragraph. Then write the numbers of each of the statements that refer to topic *A* in the space marked *A*, and those referring to topic *B* in the appropriate space. List for *X* the *numbers* of those statements that do not relate to either *A* or *B*; do not list any number more than once. For example, note that *4* should be written in the space before *B*, and *5* before *X*.

_____*A*. Name some economic influences that led to exploration and colonization.

_____*B*. Give some reasons for the English character of our laws and institutions.

_____*X*. Unrelated items.

1. Raw wealth in the form of silver and gold was eagerly sought in the New World.
2. The early colonists were given the rights of English subjects.
3. Paper money is always discounted in a foreign country.
4. The English settlers showed their interest in representative forms of government.
5. England allowed many convicted criminals to emigrate to America.
6. There was great rivalry among the nations for the control of world trade.
7. English forms and procedure were adopted by the colonial legislatures.

TEST 3, FORM A

WINNING INDEPENDENCE

1765–1789

Name.....*Grade*.....*Date*.....

I. Fill in the blank preceding each of the descriptions below with the name to which reference is made:

____ 1. The author of the Declaration of Independence.
____ 2. Our minister to France during the Revolutionary War.
____ 3. The Prime Minister of England during the Revolutionary War.
____ 4. The originator of the Committees of Correspondence.
____ 5. The attorney who defended the colonists' cause in the Writs of Assistance cases.
____ 6. The place where the first Continental Congress assembled.
____ 7. The battle which marked the turning point of the Revolutionary War.
____ 8. The commander of the British forces in the final battle of the War.
____ 9. The presiding officer of the Constitutional Convention.
____ 10. A prominent Virginian who aided in writing "The Federalist."

II. Number the following events in the order of their occurrence, or give the exact date for each. Do not combine the two forms in answering.

____ Signing the Declaration of Independence.
____ The treaty marking the French alliance.
____ Shays's rebellion.
____ The effort to enforce the Stamp Tax.
____ The battle of Saratoga.

III. Some of the following statements are true and some are false. Put a + sign before those that are true and a 0 before those that are false. Do not guess; if you do not know the answer, leave the space blank.

____ 1. Great Britain collected taxes from her subjects in England.
____ 2. The American revolutionists were undoubtedly influenced by a desire to be free from Great Britain.
____ 3. The "Critical Period" of our history refers to the winter at Valley Forge.
____ 4. France had a selfish interest in her espousal of the American cause.
____ 5. The Loyalists may have numbered as high as one third of the colonists.

____ 6. The Revolutionary War was financed by the money contributions of loyal citizens.
____ 7. The delegates to the Constitutional Convention were united as to the need for a strong, centralized government.
____ 8. The Congress under the Articles of Confederation refused to endorse the calling of a Constitutional Convention in 1787.
____ 9. Our nation faced bankruptcy after the Revolutionary War.
____ 10. The delegates to the Convention refused to deny any powers to the states.
____ 11. The Federal Supreme Court was a new feature in organized government.
____ 12. Amendments to the Constitution can originate only in Congress.
____ 13. A case of mail robbery is tried in a state court.
____ 14. Ten amendments to the Constitution were adopted by the First Congress.
____ 15. The Constitution limits freedom of speech to utterances that do not affect the security of the government.

IV. In the following statements there are references to powers specified in our Federal Constitution. Read carefully the topics *A*, *B*, etc., and then write on the proper blank the *numbers* of the statements that refer to each topic. Note the blank for non-specified powers.

_____*A*. The powers of the President.
_____*B*. The powers of Congress.
_____*C*. Combined powers of President and Senate.
_____*D*. The jurisdiction of the Supreme Court.
_____*E*. Not specified in the Constitution.

1. To borrow money.
2. To make treaties.
3. To settle controversies due to different interpretations of the meaning of the Constitution.
4. To guarantee the right to vote to all citizens.
5. To see that laws are executed.
6. To settle controversies between two states.
7. To grant pardons for offenses against the United States.
8. To provide for the annexation of new territory.
9. To regulate commerce.
10. To decide cases affecting ambassadors of foreign governments.

V. Read carefully topics *A* and *B* that follow this paragraph. Then write the *numbers* of each of the statements that refer to topic *A* in the space marked *A*, and those referring to *B* in the appropriate space. List for *X* the *numbers* of those statements that do not relate to either *A* or *B*; do not list any number more than once.

_____*A*. Conditions leading to compromises in the Constitution.
_____*B*. Conditions indicating weaknesses in the Articles of Confederation.
_____*X*. Irrelevant statements are to be listed here.

1. The large number of slaves in the Southern colonies.
2. The difficulty of making treaties with foreign governments.
3. The question of representation in the legislative branch.
4. Sectional differences among the colonies as to industrial interests.
5. The refusal of the British to give up the Northwest forts.
6. The absence of an executive branch of the government.
7. The provision of unanimous consent for amendment.
8. The effort to induce the Canadian provinces to join our group.
9. The variation in size and population in the different colonies.
10. The variation among the colonies as to issues of currency.

TEST 3, FORM B

WINNING INDEPENDENCE

1765–1789

Name.....*Grade*.....*Date*.....

I. Fill in the blank preceding each of the descriptions below with the name to which reference is made:

____ 1. The author of *Poor Richard's Almanac.*
____ 2. The king to whom the "Olive Branch Petition" was addressed.
____ 3. An English statesman who championed the cause of the colonies in the House of Commons.
____ 4. The patriot whose name was linked with that of Samuel Adams in the British orders for the arrest of the two men.
____ 5. The English general who was sent over to close the port of Boston.
____ 6. The city where the first President was inaugurated.
____ 7. The last battle of the Revolutionary War.
____ 8. A Virginian who opposed the adoption of the Constitution.
____ 9. The merchant who sacrificed his own fortune while managing the finances of the Revolution.
____ 10. A leading New Yorker who aided in writing "The Federalist."

II. Number the following events in the order of their occurrence, or give the exact date for each. Do not combine the two forms in answering.

____ The battles of Concord and Lexington were fought.
____ The Constitutional Convention met.
____ The French openly espoused the American cause.
____ Washington was inaugurated President.
____ The treaty which ended the Revolutionary War was signed.

III. Some of the following statements are true and some are false. Put a + sign before those that are true and a 0 before those that are false. Do not guess; if you do not know the answer, leave the space blank.

____ 1. The Writs of Assistance were issued to enable American ships to avoid paying English customs duties.
____ 2. Economic interests in England forced the repeal of the Colonial Stamp Tax.
____ 3. White settlements were made beyond the Alleghenies before the Revolutionary War began.
____ 4. At no time during the Revolutionary War was the outcome doubtful.

____ 5. The treaty that ended the Revolutionary War was a triumph for the United States; France gained nothing.
____ 6. The colonists were generous in their treatment of the Loyalists.
____ 7. The American victory was due to factors other than the number and equipment of military forces.
____ 8. Washington had enemies who tried to discredit him in Congress.
____ 9. During the years immediately following 1783 this nation was prosperous, at peace, and united under a stable government.
____ 10. The Constitutional Convention had representatives from all of the thirteen colonies.
____ 11. Groups of interested spectators crowded the hall where the Constitutional Convention was in progress.
____ 12. The "check and balance system" was one of the defeated compromises of the Constitutional Convention.
____ 13. The Constitution holds that the Federation has the right to deny certain powers to the states.
____ 14. A unanimous vote of the states is required to ratify an amendment to the Constitution.
____ 15. Under the Articles of Confederation our economic stability was unquestioned.

IV. The following statements contain references to powers of government. Some of these are specified in the Constitution and some are not. Read carefully the topics *A*, *B*, *C*, and *D*, and on the blank preceding each topic write the *numbers* of the statements referring to powers specified in the Constitution for that department. Note the blank for non-specified powers.

_____ *A*. Powers of the President.
_____ B. Powers of Congress.
_____ *C*. Combined powers of President and Senate.
_____ *D*. The jurisdiction of the Supreme Court.
_____ *E*. Not specified in the Constitution.

____ 1. To establish post-offices and post roads.
____ 2. To provide for copyrights and patents.
____ 3. To settle maritime disputes.
____ 4. To call extra sessions of Congress.
____ 5. To appoint ambassadors.
____ 6. To direct action when a state is a contesting party.
____ 7. To coin money and fix standards of weight and measure.
____ 8. To decide conditions under which a state may secede.
____ 9. To make treaties.
____ 10. To settle disputes involving interpretation of the Constitution.

V. Read carefully topics *A* and *B* that follow this paragraph. Then write the *numbers* of each of the statements that refer to topic *A* in the space marked *A*, and those referring to *B* in the appropriate space. List for *X* the *numbers* of those statements that do not relate to either *A* or *B*; do not list any number more than once.

_____ *A*. State some conditions which led to the Revolution.
_____ *B*. State some evidences of the Critical Period.
_____ *X*. List the numbers of unrelated statements here.

1. Legislation that was repressive and arbitrary.
2. The absence of unified, effective authority.
3. Difficulty in adjusting customs rates.
4. The desire for religious freedom.
5. Objection to a government centralized at a distance.
6. Retaliatory tariff measures between different colonies.
7. Inability to secure foreign credit.
8. Lawlessness due primarily to deflation of the currency.
9. Difficulty in negotiating trade treaties.
10. Unwillingness to contribute to the financial support of the government without adequate legislative participation.

TEST 5, FORM A

COLONIZATION AND INDEPENDENCE

1492–1836

INSTRUCTIONS TO THE STUDENT:

It is not expected that you will know all of the following facts nor that you will have time to answer all of them. Answers are to be written in the blanks on this paper. Before you begin to write be sure to read the directions at the top of the section on which you are working. Then fill in the answers which you know, without waiting to fill in the doubtful ones. Go through all the sections, filling in the answers you know best. If there is time left, you may come back to the questions which you omitted.

A "passing mark" on this test will depend on the average number of correct answers given by the pupils in your class and not upon a percentage of answers which an outsider might think you ought to know. So do your best to be one of those to get the largest number of correct answers.

Those of you who are accustomed to the fifty-point test should remember that this is just twice as long, but you are not allowed double time.

Name.....*Grade*.....*Teacher*.....
Date.....*Section*.....

I. When the following pairs of words mean the same, or nearly the same, put a + sign in the blank following each; if their meaning is opposite, put a 0 in the blank.

1. political—governmental.....
2. executive—administrative.....
3. radical—conservative.....
4. alliance—confederation.....
5. neutral—belligerent.....
6. nullify—sanction.....
7. reprisal—retaliation.....
8. ratify—abrogate.....
9. implied powers—strict construction.....
10. states' rights—Federal control.....

II. Indicate the chronological order of the events in each group by marking them 1, 2, 3, or 4. For example; 3 Civil War; 2 Washington elected; 1 America discovered.

A.
____ Jamestown settled.
____ Battle of Bunker Hill.
____ Constitutional Convention.
____ Declaration of Independence.

B.
____ The "Bloodless" Revolution.
____ Stamp Tax Congress.
____ Last of Thirteen Colonies founded.
____ Explorations of Lewis and Clark.

C. ____ Monroe Doctrine.
____ First tariff law enacted.
____ South Carolina passed Nullification Act.
____ Battle of Yorktown.

D. ____ Invention of cotton gin.
____ Erie Canal built.
____ The Clermont steamed up the Hudson.
____ The Cumberland Road completed.

E. ____ Braddock was defeated.
____ France ceded Canada to England.
____ Northwest Ordinance was adopted.
____ Spain sold Florida to the United States.

III. Some of the following statements are true and some are false. Put a + sign before those which are true and a 0 before those which are false. Do not guess; if you do not know the answer, leave the space blank.

____ 1. England gave her colonies more freedom than did France.
____ 2. The French made few explorations of importance in North America.
____ 3. The Spanish made no permanent settlement in North America.
____ 4. The Spanish flag was the first to circumnavigate the globe.
____ 5. All residents of English colonies were permitted to vote for members of Colonial legislatures.
____ 6. The Continental Congress had authority to raise money by taxation.
____ 7. The Articles of Confederation were too easily amended.
____ 8. Our nation was unable to pay its debts at the close of the Revolutionary War.
____ 9. The United States refused to pay the state debts due to the Revolution.
____ 10. The tract of land included in the Louisiana Purchase was larger than the original Thirteen Colonies.
____ 11. Each state has an equal number of representatives in Congress.
____ 12. The Southern states favored the War of 1812.
____ 13. Manufacturing developed rapidly in the South from 1815 until 1830.
____ 14. The Hartford Convention discussed secession.
____ 15. The national government helped to build the Cumberland Road.
____ 16. The Embargo Act helped commerce in New England.
____ 17. The Missouri Compromise was popular in the South after 1840.
____ 18. Many social reform movements began in the 1830's.
____ 19. The Monroe Doctrine planned to offset England's activity in South America.
____ 20. Jefferson's policies deviated from the Democratic-Republican platform.

IV. Fill in the blank preceding each description with the name of the person to whom reference is made.

..... 1. The English king who announced the "divine right" theory.

..... 2. The first "English sea-rover" to circumnavigate the globe.

..... 3. The French explorer who was the first governor of Canada.

..... 4. The American statesman who suggested the Albany Plan of Union.

..... 5. The Prime Minister of England during the Revolutionary War.

..... 6. The New England patriot who organized the Committees of Correspondence.

..... 7. The Virginian who drafted the Declaration of Independence.

..... 8. The inventor of the cotton gin.

..... 9. The president of the Constitutional Convention.

.....10. The Secretary of the Treasury under Washington.

.....11. The builder of the first practical steamboat.

.....12. The Englishman who invented the steam locomotive.

.....13. An explorer whom Jefferson sent to the Far West (Any one of three.)

.....14. The woman who did much to improve the treatment of the insane.

.....15. The early Chief Justice whose important decisions are largely responsible for the prestige of the Supreme Court.

.....16. A Southern statesman and Whig known for his advocacy of compromise legislation.

.....17. The author of the "Exposition and Protest."

.....18. The Northern senator who defended the Union in the debate with Hayne.

.....19. An anti-slavery agitator and editor of "The Liberator."

.....20. The man who introduced the "spoils system" into presidential politics.

V. Write the word in each of the blanks below which will make the statement true with reference to the Constitution of the United States.

1–2. The President's veto can be overridden by a .....vote of .....

3. Bills of revenue must originate in the .....
4. The President must be a.....citizen.
5. Treaties must be ratified by the .....
6. The.....has power to declare acts of Congress unconstitutional.
7. Congress shall meet at least .....

8–10. Amendments to the Constitution are usually proposed by..... and ratified by a vote of..... of the .....

VI. Write the numbers of each of the statements below which refer to Topic *A* in the space marked *A*; then list those for *B* and for *C*. For *X* list the numbers of those statements which do not relate to either *A*, *B*, or *C*; list no number more than once. As an example of a correct answer, number 16 is listed after *A*.

A. Conditions leading to the Revolutionary War.
B. "Loose construction" of the Constitution.
C. Jacksonian Democracy.
X. Irrelevant statements.

A. *16*_____
B. _____
C. _____
X. _____

1. The government was centralized at a distance.
2. The doctrine of "implied powers" was accepted.
3. There was dissatisfaction over customs rates.
4. Opposition was expressed to a government grant of financial monopoly.
5. The invention of the cotton gin was important.
6. In the reaction from conservatism, manhood suffrage became popular.
7. Much repressive legislation was enacted.
8. Federal soldiers were used to enforce a high tariff.
9. A treaty of alliance was made with France.
10. There was resistance to supporting a government which refused legislative participation.
11. Supreme Court decisions established a judicial precedent.
12. The charter of the First National Bank was thus made possible.
13. The opponents declared the Federal government was the agent of the states.
14. Political awards for successful party adherents were advocated.
15. The purchase of Louisiana was an example of its operation.
16. George III was stubborn and arbitrary.

VII. The study of history should include a knowledge of where to look for desired information. In the blanks below list any author for each of the various kinds of reference works asked for. If you do not know the author, give a title which identifies the book.

.....1. Author or editor of a *source* book of American history.

.....2. The author of *The Critical Period of American History.*

.....3. The author of an encyclopedia of historical facts.

.....4. Designate an authoritative series of volumes dealing with the whole range of American history.

.....5. Give author and title of any volume dealing with a single phase of American history; do not repeat one given above.

VIII. The following questions are designed to show the relationship and similarity of certain men and events to other men and events. Select from the words given in the parentheses the one which best expresses the relation; write this in the proper blank.

Example: *Franklin* was to *Poor Richard's Almanac* what *Hamilton* was to.....
(Madison—New York—Federalist—Union)
(The word *Federalist* should be written in the blank.)

1. *Taxation* without representation was to the *Colonist* what *rebellion* was to.....
(Massachusetts—France—smuggling—Loyalists)
2. *Washington* was to the *Revolutionary War* what *Franklin* was to the .....
(French Alliance—electricity—printing—Constitution)
3. *England* was to the *Monroe Doctrine* what *Jefferson* was to the .....
(Federalists—War of 1812—Louisiana Purchase—high tariff)
4. *Paper money* was to the *Continental Congress* what the *French Alliance* was to .....
(Washington—LaFayette—Valley Forge—foreign tariff)
5. *Strict Construction* was to the *Federalists* what..... was to *Jackson.*
(spoils system—Purchase of Florida—Democrat—National Bank)

SCORING KEY FOR TEST 2, FORM A

DISCOVERY AND EXPLORATION

1492–1763

I. 1. Africa
2. Canada (North America)
3. America (North or South)
4. South America
5. North America (Pennsylvania)

II. 1. Waldseemüller
2. Cabral
3. Toscanelli
4. Champlain
5. James I
6. Wolfe
7. Gilbert
8. Coronado
9. Drake
10. Bacon
11. Hutchinson
12. Locke
13. Braddock
14. Andros
15. Franklin

III. 1620 — 3
1603 — 2
1689 — 4
1521 — 1
1754 — 5

IV. 1. 0
2. 0
3. +
4. 0
5. +
6. 0
7. +
8. +
9. 0
10. 0
11. +
12. 0
13. 0
14. +
15. +
16. +
17. 0
18. 0
19. +
20. 0

V. (1. — 9) (Sample)
2. — 11
3. — 7
4. — 10
5. — 12
6. — 8

(*Total:* 50 points; multiply by 2 to get percentage of 100.)

SCORING KEY FOR TEST 2, FORM B

DISCOVERY AND EXPLORATION

1492–1763

I. 1. Africa
2. Canada (North America)
3. South America
4. Africa
5. North America (Pennsylvania)

II. 1. Raleigh
2. Charles I
3. Winthrop
4. Cabot
5. Prince Henry
6. Magellan
7. Cartier
8. La Salle
9. Hooker
10. Berkeley
11. Washington
12. Stuyvesant
13. Montcalm
14. Pitt
15. Fox

III. 1607 — 3
1763 — 5
1588 — 2
1497 — 1
1619 — 4

IV. 1. 0
2. +
3. +
4. +
5. 0
6. 0
7. +
8. 0
9. 0
10. +
11. +
12. 0
13. +
14. 0
15. 0
16. +
17. +
18. 0
19. +
20. 0

V. *A*. 1 — 6
B. 2 — 7
X. 3

(Answers 4 and 5 are used as illustrations.)

(*Total:* 50 points; multiply by 2 to get percentage of 100.)

SCORING KEY FOR TEST 3, FORM A

WINNING INDEPENDENCE

1765–1789

I. 1. Jefferson
2. Franklin
3. Lord North
4. Samuel Adams
5. Otis
6. Philadelphia
7. Saratoga
8. Cornwallis
9. Washington
10. Madison

II. 1776 — 2
1778 — 4
1786 — 5
1765 — 1
1777 — 3

III.

1. +	6. 0	11. +
2. +	7. +	12. 0
3. 0	8. 0	13. 0
4. +	9. +	14. +
5. +	10. 0	15. 0

IV. *A.* 5 — 7
B. 1 — 9
C. 2
D. 3 — 6 — 10
E. 4 — 8

V. *A.* 1 — 3 — 4 — 9
B. 2 — 6 — 7 — 10
X. 5 — 8

(*Total:* 50 points; multiply by 2 to get percentage of 100.)

SCORING KEY FOR TEST 3, FORM B

WINNING INDEPENDENCE

1765–1789

I. 1. Franklin
2. George III
3. Pitt or Burke
4. Hancock
5. Gage
6. New York
7. Yorktown
8. Patrick Henry or Lee
9. Robert Morris
10. Hamilton or Jay

II. 1775 — 1
1787 — 4
1778 — 2
1789 — 5
1783 — 3

III.

1. 0	6. 0	11. 0
2. +	7. +	12. 0
3. +	8. +	13. +
4. 0	9. 0	14. 0
5. +	10. 0	15. 0

IV. *A*. 4
B. 1 — 2 — 7
C. 5 — 9
D. 3 — 6 — 10
E. 8

V. *A*. 1 — 3 — 5 — 10
B. 2 — 6 — 7 — 8 — 9
X. 4

(*Total:* 50 points; multiply by 2 to get percentage of 100.)

SCORING KEY FOR TEST 5, FORM A

COLONIZATION AND INDEPENDENCE

1492–1836

I. 1. +
2. +
3. 0
4. +
5. 0
6. 0
7. +
8. 0
9. 0
10. 0

II. *A.* 1 — 2 — 4 — 3
B. 1 — 3 — 2 — 4
C. 3 — 2 — 4 — 1
D. 1 — 3 — 2 — 4
E. 1 — 2 — 3 — 4

III. 1. +
2. 0
3. 0
4. +
5. 0
6. 0
7. 0
8. +
9. 0
10. +
11. 0
12. +
13. 0
14. +
15. +
16. 0
17. 0
18. +
19. 0
20. +

IV. 1. James I
2. Drake
3. Champlain
4. Franklin
5. Lord North
6. Samuel Adams
7. Jefferson
8. Whitney
9. Washington
10. Hamilton
11. Fulton
12. Stephenson
13. Lewis-Clarke-Pike
14. Dix
15. Marshall
16. Clay
17. Calhoun
18. Webster
19. Garrison
20. Jackson

V. 1–2. $\frac{2}{3}$ — Congress (Both houses) [1]
3. House of Representatives
4. native-born
5. Senate
6. Supreme Court
7. once in every year
8–10. Congress — $\frac{3}{4}$ — of state legislatures

VI. *A*. 1 — 3 — 7 — 10 — 16
B. 2 — 11 — 12 — 13 — 15
C. 4 — 6 — 8 — 14 (3 — 13) [1]
X. 5 — 9

VII. 1. Hart — MacDonald, etc.
2. Fiske
3. Larned — Harper, etc.
4. Yale *Chronicles; American Nation Series*, etc.
5. Thwaite: *Colonies;* Earle: *Home Life in Colonial Days*, etc.

VIII. 1. Loyalists
2. French Alliance
3. Louisiana Purchase
4. Washington
5. National Bank

(*Total:* 100 points.)

[1] Duplicate answers given in parentheses may be accepted as correct.

BIBLIOGRAPHY

PUBLISHED TESTS IN THE SOCIAL STUDIES

HISTORY

American Council European History Test. By H. J. Carman, W. C. Langsam, and B. D. Wood. World Book Company, Yonkers-on-Hudson, New York; 1928.

For high school and college. Consists of well-selected questions sampling a wide range of material. True-false, matching, multiple-choice, and recall types of questions are used. Entirely objective. Standardized. Alternate forms. 145 items.

Background Tests in Social Science. By Tyler Kepner. Ginn & Co., Boston; 1924.

For high school. Covers material selected from European and American history, and civics. Multiple-choice, recall, and matching types of questions are used. Entirely objective. Standardized. Alternate forms. 100 items.

Barr Diagnostic Tests. By A. S. Barr. Public School Publishing Company, Bloomington, Illinois; 1920.

For Grades 8 to 12. The test is informational rather than diagnostic, although the items do not cover the entire range of American history. Consists of varied material valuable for experimental purposes rather than for measurement. Multiple-choice and matching types of questions are used. Not entirely objective. Cannot be scored quickly. Two forms. About 150 items.

Columbia Research Bureau American History Test. By H. J. Carman, T. N. Barrows, and B. D. Wood. World Book Company, Yonkers-on-Hudson, New York; 1926.

For high school and college. Well-selected material including political, economic, and social aspects of history. True-false, matching, and multiple-choice types of questions are used. Entirely objective. Standardized. Alternate forms. 200 items.

Denny-Nelson American History Test. By E. C. Denny and M. J. Nelson. World Book Company, Yonkers-on-Hudson, New York; 1928.

For Grades 7 and 8. Consists of a good sampling of material of varied range. Divided into Part I, Pre-National Period, including 80 items; and Part II, National Period, including 100 items. The booklet is printed in two parts which may be used at different times. Multiple-choice, matching, recall, true-false, and yes-no types of questions are used. Entirely objective. Standardized. Alternate forms. 180 items.

Gregory Tests in American History. By C. A. Gregory. Bureau of Administrative Research, University of Cincinnati; 1923.

There are separate tests for seventh and eighth grades. Test III for

Grades 8 to 12 includes material that is largely political. Recall and multiple-choice types of questions are used. Entirely objective. Standardized. Alternate forms. 100 items.

Iowa General Information Test in American History. By M. H. DeGraff, G. M. Ruch, and H. A. Greene. Bureau of Educational Research, University of Iowa, Iowa City; 1927.

For Grades 7 to 12. Each of the two forms contains 100 items selected from a study of three widely used textbooks. Recall type of questions is used. Entirely objective. Standardized. Alternate forms.

Junior American History Test. By H. J. Carman, T. N. Barrows, and B. D. Wood. World Book Company, Yonkers-on-Hudson, New York; 1929.

For junior high school. Careful sampling of various types of material. True-false, matching, and multiple-response types of questions are used. Entirely objective. Standardized. Alternate forms. 135 items.

Lynch American History Test Series. By I. R. Lynch. Educational Test Bureau, Minneapolis; 1926.

For junior high school. A series of tests adapted to each of six periods of American history. Questions cover material usually found in eighth-grade tests. Recall, multiple-choice, true-false, and matching types of questions are used. Entirely objective. One form. 75 items.

New Stanford Achievement Test. By T. L. Kelley, G. M. Ruch, and L. M. Terman. World Book Company, Yonkers-on-Hudson, New York; 1929.

For Grades 4 to 9. Ten subject tests are included in this group of achievement tests. 80 items are included in the History and Civics Test, and 80 items in the Geography Test. The selection of material is well balanced. Multiple-choice type of questions is used. Entirely objective. Standardized. Five forms.

Public School Achievement Test in History. By J. S. Orleans. Public School Publishing Company, Bloomington, Illinois; 1928.

For Grades 5 to 8. The test is elementary in character and includes political material. Multiple-choice, matching, and completion types of questions are used. Entirely objective. Alternate forms. 68 items.

Understanding of American History. By L. W. Pressey and R. C. Richards. Public School Publishing Company, Bloomington, Illinois.

For Grades 6 to 12. This is largely a test of information rather than understanding, as the title would seem to indicate. Material covered is political. 75 information items and 25 items requiring the matching of descriptive adjectives to the names of statesmen. Multiple-choice type of questions is used. Largely objective. Alternate forms. 100 items.

Van Wagenen American History Scales (Revised). By M. J. Van Wagenen. Bureau of Publications, Teachers College, Columbia University, New York.

For Grades 5 to 8. Material is well balanced but selected from too brief a sampling. Multiple-choice type of questions is used. Largely objective. Duplicate forms. 30 items.

Witham Comprehensive History Test. By E. C. Witham. J. L. Hammett Company, Newark, New Jersey.

For Grades 7 to 8. Material is covered by means of completion exercises. Objective. Standardized. One form. 30 items.

CIVICS

American Council Civics and Government Test. By T. F. Leigh, J. D. McGoldrick, P. H. Odegard, and B. D. Wood. World Book Company, Yonkers-on-Hudson, New York; 1929.

For high school and college. Significant selection covering a wide range of material in the field of political science. True-false, matching, multiple-choice, and completion types of questions are used. Entirely objective. Standardized. Alternate forms. 222 items.

Brown-Woody Civics Test. By A. W. Brown and C. Woody. World Book Company, Yonkers-on-Hudson, New York; 1928.

For Grades 7 to 12. Designed for high schools but may also be used for lower grades. Tests civic vocabulary, information, and judgment. Entirely objective. Standardized. Alternate forms. 170 items.

Burton Civics Test. By W. H. and V. N. Burton. World Book Company, Yonkers-on-Hudson, New York; 1928.

For Grades 5 to 9. Material tests civic information, with emphasis on understanding of terms rather than memorized detail. Multiple-choice type of questions is used. Entirely objective. Standardized. Alternate forms. 60 items.

Hill Civics Test. By H. C. Hill. Public School Publishing Company, Bloomington, Illinois; 1926.

For Grades 6 to 12. Includes two separate tests, one in Civic Attitudes and one in Civic Information. Multiple-choice type of responses is used. The information questions are objective. One form. 40 items.

ECONOMICS

American Council Economics Test. By H. Taylor, T. N. Barrows, and B. D. Wood. World Book Company, Yonkers-on-Hudson, New York; 1929.

For high school and college. Covers a wide sampling of carefully selected material. True-false, matching, and multiple-choice types of questions are used. Entirely objective. Standardized. Alternate forms. 223 items.

GEOGRAPHY

Buckingham-Stevenson Place Geography Tests. By B. R. Buckingham and P. R. Stevenson. Public School Publishing Company, Bloomington, Illinois.
For Grades 5 to 8. Material covers world and United States geography in two separate tests. Dictation exercise requiring location of places. Fairly objective. Three forms. 130 items.

Gregory-Haggerty Geography Tests. By C. A. Gregory and M. E. Haggerty. Bureau of Administrative Research, University of Cincinnati.
For Grades 4 to 6. Material includes five parts covering facts on world geography. Miscellaneous questions, location and map studies. Various types of questions are used. Fairly objective. Two forms. 100 items.

Gregory-Spencer Geography Tests. By C. A. Gregory and P. L. Spencer. Bureau of Administrative Research, University of Cincinnati.
For Grades 6 to 8. Material divided into eight parts testing knowledge of trade routes and products, miscellaneous geography, causal geography of United States and of world, place and descriptive geography, and map studies. Various types of questions are used. Objective. Three forms. 100 items.

New Stanford Geography Test. By T. L. Kelley, G. M. Ruch, and L. M. Terman. World Book Company, Yonkers-on-Hudson, New York; 1929.
For Grades 4 to 9. Material covers world geography. Multiple-choice type of questions is used. Entirely objective. Standardized. Five forms. 80 items.

Public School Achievement Test in Geography. By J. S. Orleans. Public School Publishing Company, Bloomington, Illinois; 1928.
For Grades 4 to 8. Material is divided into four parts and includes miscellaneous questions on world geography and map questions. Completion, true-false, and multiple-choice types of questions are used. Objective. Two forms. 62 items.

Stevenson-Ridgley-Shipman Geography Test. By P. R. Stevenson, D. C. Ridgley, and J. M. Shipman. Public School Publishing Company, Bloomington, Illinois.
Grades according to content. Material includes four tests: one for the United States, one for South America, one for Europe, and one for Asia. Each test includes one part on information and one on problems. Multiple-choice type of questions is used. Entirely objective. Duplicate forms. 75 items.

PUBLISHED READING TESTS

Detroit Reading Test. By C. M. Parker and E. A. Waterbury. World Book Company, Yonkers-on-Hudson, New York.
For Grades 2 to 9. Material includes four tests, each for different grades,

testing reading comprehension; and five for measuring progress from term to term. Objective. Two forms. 12 paragraphs.

Haggerty Reading Examination, Sigma 3. By M. E. and L. C. Haggerty. World Book Company, Yonkers-on-Hudson, New York.
For Grades 6 to 12. Material covers vocabulary, sentence reading, and paragraph reading. Various types of questions are used. Objective. Two forms.

New Stanford Reading Test. By T. L. Kelley, G. M. Ruch, and L. M. Terman. World Book Company, Yonkers-on-Hudson, New York; 1929.
For Grades 2 to 9. Material includes tests in paragraph meaning and word meaning. Completion and multiple-choice types of responses are used. Objective. Standardized. Five forms.

Public School Achievement Test in Reading. By J. S. Orleans. Public School Publishing Company, Bloomington, Illinois; 1928.
For Grades 2 to 8. Reading exercise covering a variety of material. Various types of questions are used. Objective. Two forms.

Sangren-Woody Reading Test. By P. V. Sangren and C. Woody. World Book Company, Yonkers-on-Hudson, New York.
For Grades 4 to 8. Material is divided into seven parts, which test Word Meaning, Rate, Fact Material, Total Paragraph Meaning, Central Thought, Following Directions, and Organization. Various types of questions are used. Objective. Two forms.

Stone Narrative Reading Tests. By C. R. Stone, L. C. Murphy, and A. Buehrmann. Public School Publishing Company, Bloomington, Illinois.
For Grades 3 to 9. Material includes three separate story books, including five stories. Tests are on material in stories and rate of grading. Multiple-choice type of questions is used. Objective. One form.

Thorndike-McCall Reading Scale. By E. L. Thorndike and W. A. McCall. Bureau of Publications, Teachers College, Columbia University, New York.
For Grades 2 to 12. Each test contains eleven paragraphs with thirty-five questions on each paragraph. Objective. Ten equivalent forms.

Van Wagenen Reading Scales. By M. J. Van Wagenen. Public School Publishing Company, Bloomington, Illinois.
For Grades 8 to 12. Material includes two tests, one for History and General Science, the other for Literature. Each test contains fifteen paragraphs with statements to be checked to test the pupil's comprehension of material. Objective. Duplicate forms.

Whipple High School and College Reading Test. By G. M. Whipple. Public School Publishing Company, Bloomington, Illinois.
For high school and college. Material consists of six pages of reading matter with questions and directions which the reader is required to answer. Objective. Two forms.

PUPIL WORKBOOKS AND STUDY GUIDES[1]

American History Workbook. By M. C. Bishop and E. K. Robinson. Ginn & Co., Boston; 1929.
May be used "to supplement any of the leading textbooks in American history used in the upper elementary and junior high school grades." Includes review outlines, maps and map exercises, suggestions for stories and essays, pictures, and sixteen detachable tests. 111 pages.

California. By Harold Hughes. Ginn & Co., Boston; 1929.
Includes descriptive material, study exercises, maps, and tests. 72 pages.

Directed History Study, Book One (1928), Book Two (1926), Book Three (1927). By C. C. Scheck and M. A. Orton. World Book Company, Yonkers-on-Hudson, New York.
For sixth, seventh, and eighth grades. Include page assignments in various history texts for each of the problems stated, questions, map exercises, and supplementary reading references. 96, 123, and 123 pages respectively.

Directed Study Book for World History. By C. Perkins. Rand McNally & Co., Chicago; 1929. 70 pages.

Directive Study Sheets in World History for High School Students. By C. C. Barnes. Houghton Mifflin Company, Boston; 1929.
For each unit an outline is provided with reading references and assignments for study for special reports and for activities. 142 pages.

Laboratory Manual in American History. By H. E. Wilson. American Book Company, New York; 1927.
A syllabus for high school history including a topical outline for each unit, with references to various texts and suggestions for additional topical reading, as well as questions, problems, maps, map exercises, and tests. This book is especially valuable for the reading lists given for each unit. These lists cover not only collateral references, but a second grouping of biography, travel, and reminiscence, and a third group of imaginative literature. 241 pages.

Mace American History Notebook. By C. B. Ulery. Rand McNally & Co., Chicago; 1929.
Designed to accompany a particular text; thought questions, map exercises, and names to be identified are included. 99 pages.

My Community. By H. C. Hill and D. H. Sellers. Ginn & Co., Boston; 1927.
Includes descriptive material, study exercises, some statistical tables, graph sheets, and blank pages. 147 pages.

Outline Workbook for American History. By N. M. Holmes. Henry Holt & Co., New York; 1929.
Includes page references to different texts, spaces for completing details

[1] Neither this list nor those given previously attempt to include all published material. Certain representative titles have been included to meet different needs.

in outlines on leading topics, supplementary exercises, topical reviews, sample test questions, outline maps, and a "Minimum Reference Library" list.

Pupil's Workbook in Geography of California. By F. A. Rice and W. G. Paden. Ginn & Co., Boston; 1922.
A workbook based on the problem method. Material includes a liberal supply of maps, statistical tables, questions, and graphs to be completed. 81 pages.

Pupil's Workbook to Accompany "Changing Civilization in the Modern World." By H. Rugg and J. Mendenhall. Ginn & Co., Boston; 1929.
Includes suggestions for study, discussion, debates, excursions, graphs, and maps, with a brief test for each lesson. 80 pages. (There is a similar workbook to accompany Rugg's *An Introduction to American Civilization.*)

Student's Objective Test Manual. By H. C. Perkins. Ginn & Co., Boston; 1930.
This manual is designed to accompany Muzzey's *History of the American People* and includes questions and short objective tests adapted to certain numbered paragraphs. 328 pages.

Study Guide Tests in American History. By M. J. Stormzand. Macmillan Company, New York; 1928.
This manual is designed to accompany Beard's school history and includes questions for each lesson phrased in objective form. 64 pages.

Twenty Tests. By M. C. Bishop and E. K. Robinson. Ginn & Co., Boston; 1929.
Includes tests taken from *Map Exercises, Syllabus, and Notebook* and designed to accompany Muzzey's *History of the American People.* 37 pages.

Workbook in United States History for Use with "The Growth of a Nation." By W. P. Webb. Row, Peterson & Co., Evanston, Illinois; 1929.
142 pages.

Work-Test Manual in American History. By W. A. Butcher. Macmillan Company, New York; 1929.
Includes questions to be completed, assignments, and brief references. 192 pages.

INDEX

Ability, variations in, 48, 57
Accuracy, historical, 7, 55
Achievement in history, measure of, 59–60, 131–133; scaling, 82; of average student, 131–135
Administering tests, 112–113
Alternate forms, construction of, 101–104
Ambiguity in questions, 111
American Historical Association, plans for test survey, 139–140
Appearance of completed test, 114–116
Attitudes, as aim, 12, 14, 26; relation to experiences, 13; tests of, 90–91

Background tests, 46–47
Barr Diagnostic Tests, 46
Best-answer form, 66

Chance, correction for, 78–79
Columbia Research Bureau Tests, 47, 86, 104
Completion question form, 64, 105
Comprehension, reading, as diagnosis, 41; significance of, 42–44
Comprehensiveness of objective test, 33–35
Conditioned response in history, 19, 21–23
Construction of examinations, deserving thought, 7, 31; technique, 63–75, 99–112; difficulties to avoid, 109–112; coöperative, 85, 93, 103
Content of examination, in relation to form, 104, 119. *See also* Material
Controversial topics, 6–7, 32, 52, 67, 69, 85, 89, 97
Correlation, as index of reliability, 35
Criticism, preliminary, of test questions, 85, 103

Deliberation and time limit, 81–82
Dewey on reflective thinking, 17–18, 21, 26, 81
Diagnosis of learning difficulties, 41–46
Diagnostic tests, 42–47, 131
Difficulties in test construction, 109–112
Directions for administering tests, 112–114
Discussion, dangers of, 2–3; values of, 98; of test results, 96–97
Distribution, influenced by comprehensiveness, 35; explanation for students, 97

Epochs, emphasis on various, 104
Errors, in form, 111–112; in scoring, 118
Essay-response, 32, 37, 57–58
Examinations, purpose of, 30–32, 41, 48; differences in new-type and essay-response, 33–34; forms of new-type, 63–75. *See also* Achievement, Comprehensiveness, Diagnosis, Essay-response, Reliability, Validity
Experience as factor in learning, 12–16
Experiments with objective tests, 37, 43, 93, 121, 140

Facts, overemphasis on, 3; in relation to problem, 5–6, 24–25; in relation to reasoning, 16–18. *See also* Information
Forms. *See* Question forms
Formula, correction, 79

Goal, student's, 54, 56
Guessing in examinations, 78–79

Honesty in examinations, safeguarding, 116

Ideals, difficulty in testing, 85–86; not measured by information, 138
Individual differences, diagnosis of, 41–42; in abilities, 57
Individualized instruction, 48–53
Information, needed for critical thinking, 14–18, 20–22, 24–27, 55–56; tests of, 8–9, 39, 99–100
Inverted question, 108–109
Iowa study, 37, 61, 76–77, 79, 83
IQ, as individual measure, 48; in experimental schools, 130; influence on achievement, 133–136
Items, number of, 83, 112; careful selection of, 99

Kepner Background Test, 46

Learning history, psychology of, 11–28; diagnosis of difficulties in, 41–46

Margins, used for responses, 115–116
Marks, on examinations, 32, 37, 56–59, 94–95; improving, 60–61; varying standards for, 123–124; relation to test scores, 125–129
Material, assembling, 99; proportioning, 101–104; poor selection of, 109–110; form subordinate to, 119. *See also* Content
Median score, explanation for students, 97; in experimental tests, 129, 135, 137
Mimeographing completed test, 114–115
Minimum requirements, equalizing, 122–124
Modern Language Association, evidence of test reliability, 36
Motivation through self-measurement, 56–58
Multiple-choice form, 65–69, 85–86; recognition form, 65; best answer, 66; topical-selection, 67; content adapted to, 106–108

New-type question. *See* Examination, Objective tests, Question forms

Objective tests, function of, 6–8; characteristics, 30–39; scoring, 77–82, 117–118; in series, 93; for practice, 95
Open-book test, 31, 52–53

Percentile rank, 61
Practice tests, 95–96
Problem, as method, 2, 5–6, 21, 23, 25, 52–53
Progressive education, 4–5, 7, 11–12, 25, 139

Quartile, explanation for students, 97
Question forms, new-type, 63–75; distinctive qualities, 63, 66–67, 72–74, 105; adapting, to content, 104–108; subordinate to content, 119

Questions, good, 108; poor, 109–111

Range, in relation to comprehensiveness, 35; explanation for students, 97
Rank method, 60–61
Reading records, checking, 51
Reading tests, standard, 41–42; informal, 43, 95–96; in Rochester study, 43–44
Reasoning, in social studies, 9, 16–27; questions, 84; questions illustrated, 86–90
Recall questions, 63–64, 104–105
Recognition questions, 63, 65
Reliability, of objective test, 35–36; of old-type, 37; of various forms, 75–76
Rewarding effort, 58
Rights-minus-wrongs. *See* True-false scoring
Rochester study, 43–44, 47, 84, 87, 93, 100

Sample questions, relation of problem to fact, 5–6, 19–20; working skill, 45; open-book, 52–53; recall, 64; completion, 64; recognition, 65; three-response, 66; dual answer for controversial matter, 67; topical-selection, 68; poor true-false, 72; varied form, 74, 76; time sequence, 74–75; recognizing right answer, 84; reasoning, 86–89; inverted, 109; implied answer, 110; ambiguous, 111; informal tests, 145–161
Scaling performance, 82, 99
Scoring objective tests, 77–82, 117–118
Self-measurement, 56
Standard, student's, 53; variance among teachers, 123–124
Standardized tests, in social studies, 6–7; in reading, 41–42
Statistical terms, explanation for students, 97
Study habits, 48–50, 95
Subjective judgments, in grading tests, 37; disguised as objective response, 85; in marking students, 123–124

Teaching of history, complex skills in, 1–4; aim of, 11; psychology of, 11–28; term plan for, 93
Thinking in relation to learning, 2–3, 9, 12, 17–28, 81–82
Time, relation to reliability, 36, 76–77; limit in tests, 81–82; items possible in given, 83; amount used in testing, 98
Time sequence, form for testing, 74–75; scoring, 75
Topical-selection form, 67–68, 107
True-false form, 69–72; scoring, 77–80; content adapted to, 106

Uniform tests, value of, 122–124

Validity of objective tests, 37–39
Vocabulary tests, 42
Voters, training for, 26–27

Weighted scores, 80
Wilson's Laboratory Manual, 94
Wood, on reflective thinking, 17; on subtleties in questions, 73; on true-false content, 106; on inverted question, 109
Working Skills Test, 44–46
Work-plan, 93–94

K. S. T. C. LIBRARY

Date Due

MAR 01	1931		
MAR 10	1931		
MAY 21	1931		
FEB 12	1932		
JUL 22	1932		
JUN 28	1933		
JAN 17	1934		
JAN 2	1935		
FEB 25	1935		
MAR 11	1935		
JUL 8	1947		
DE 16 '52			
JUL 24	1953		

13-2148

WITHDRAWN

907 24434
M623
Michell, Elene
AUTHOR
Teaching values in new-type his-
TITLE
tory tests.

907
M623
Michell, Elene 24434
Teaching values in
new-type history tests.
[redacted]
RY